

# MASTER'S THESIS

## Gegevensbeschikbaarheid en -kwaliteit voor Data Analytics

Kistemaker, R.

**Award date:**  
2021

[Link to publication](#)

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



# Gegevensbeschikbaarheid en-kwaliteit voor Data Analytics.

## Data availability and quality for Data Analytics.

Opleiding:	Open Universiteit, faculteit Management, Science & Technology Masteropleiding Business Process Management & IT
Programma:	Open University of the Netherlands, faculty of Management, Science & Technology Master Business Process Management & IT
Cursus:	IM0602 Voorbereiden Afstuderen BPMIT IM9806 Afstudeeropdracht Business Process Management and IT
Student:	Rian Kistemaker
Identiteitsnummer:	
Datum:	Juli 2021
Afstudeerbegeleider	Prof.dr.ir. Remko Helms
Meelezer	Dr. Laury Bollen
Derde beoordelaar	<indien aanwezig>
Versie nummer:	v1.0
Status:	Definitief

## Abstract

Dit onderzoek gaat over het gebruik van maturity modellen. Maturity modellen (volwassenheidsmodellen) zijn geïntroduceerd als referentiekaders voor het beheer van informatiesystemen (hierna: IS) in organisaties binnen verschillende industrieën. Uit bestaand onderzoek is gebleken dat er een tekort is aan gevalideerde maturity modellen. Daarnaast is gebleken dat het gebruik van maturity modellen gericht op Data Analytics binnen het midden- en kleinbedrijf (hierna: het MKB) achter blijft in vergelijking met het gebruik daarvan binnen grote organisaties. De toename van de hoeveelheid gegevens die verzameld wordt binnen MKB-organisaties maakt het steeds moeilijker om deze te verwerken met het oog op het verbeteren van de eigen concurrentiepositie. Aangenomen wordt dat een zorgvuldig gebruik van Data Analytics bij MKB-organisaties deze bedrijven helpt om de verzamelde gegevens om te zetten in kennis, die de bedrijfsresultaten en de operationele efficiëntie kan verbeteren.

Belangrijke basiselementen bij het toepassen van Data Analytics zijn onder andere. gegevensbeschikbaarheid en gegevenskwaliteit. Dit onderzoek presenteert een gevalideerd maturity model dat gericht is op gegevensbeschikbaarheid en gegevenskwaliteit binnen MKB-organisaties en op de ondersteuning van Data Analytics. Dit model is ontwikkeld met behulp van het Design Science Research Model. Het artefact is door middel van iteraties verbeterd. Vervolgens is het verbeterde model via een case study voorgelegd aan een gebruiker in de praktijk. Resultaat is een valide model dat met name geschikt is voor gebruik door consultants die zich bezighouden met advisering van MKB-organisaties.

## Sleutelbegrippen

Focus Area Maturity model, MKB, Data Analytics, Gegevensbeschikbaarheid en Gegevenskwaliteit.

## Samenvatting

In dit onderzoek is gepoogd de vraag te beantwoorden of een maturity model ontwikkeld kan worden voor MKB-organisaties ter ondersteuning van Data Analytics. Bij het zoeken naar een antwoord is om te beginnen literatuur geraadpleegd voor een eerste verdieping in maturity modellen.

Bij het doornemen van de literatuur werd duidelijk dat maturity modellen in de praktijk minder worden toegepast binnen MKB-organisaties. Het ontbreken van specifieke kennis op dit terrein en de beperkte middelen die deze bedrijven hiervoor tot hun beschikking hebben worden als de twee belangrijkste oorzaken genoemd. Daarnaast zou ook het type maturity model dat wordt gebruikt van invloed kunnen zijn.

Er zijn drie belangrijke typen maturity modellen, met elk zijn eigen voor- en nadelen. Dat zijn het Focus Area model, het Staged 5-level model en het Continuous 5-level model. In dit onderzoek is op basis van de literatuurstudie gekozen voor het type Focus Area Maturity model. Dat heeft hier de voorkeur omdat dit model het minst aan kennis vraagt van de gebruiker en omdat het bovendien geen onnodige verplichtingen oplegt waaraan voldaan moet worden. Deze kenmerken maken het model meer toegankelijk voor het MKB, waarmee niet gezegd is dat het model daarmee minder geschikt zou zijn voor grotere organisaties. De laagdrempeligheid maakt het extra aantrekkelijk voor de kleinere onderneming

Het ontwikkelde Focus Area Maturity model, gebaseerd op het Design Science Research model, richt zich op Data Analytics. Uit literatuur onderzoek is gebleken dat gegevensbeschikbaarheid en gegevenskwaliteit belangrijke datakenmerken zijn. Met deze datakenmerken als focusgebieden in het model zijn in dit onderzoek per focusgebied telkens meerdere kenmerken beschreven. Daarnaast zijn enkele verbeterpunten beschreven welke ondersteuning kunnen bieden wanneer een bedrijf de volwassenheid van een van deze kenmerken zou willen verhogen.

Het in dit onderzoek ontwikkelde model is gevalideerd via een single case study. Op basis daarvan kan worden geconcludeerd dat het model geschikt is voor toepassing binnen een MKB-organisatie. Consultants kunnen het model gebruiken ter ondersteuning van hun advisering aan een MKB-organisatie. Zelfstandig gebruik van het model door MKB-organisaties is beperkt mogelijk. Dat vereist specialistische kennis, die vanwege de beperkte omvang en het kleine personeelsbestand van de onderneming meestal niet voorhanden zal zijn.

De resultaten van dit onderzoek kunnen in de eerste plaats van nut zijn voor MKB-organisaties bij het bepalen van de datavolwassenheid en vervolgens bij het verbeteren daarvan. Die verbetering kan ten goede komen aan de operationele efficiency, de bedrijfsresultaten en de concurrentiepositie van het bedrijf. In de tweede plaats dragen de resultaten van dit onderzoek bij aan de algemene discussie binnen Informatie Systeem (IS) groepen over maturity modellen in het algemeen.

Hoewel er op grond van de case study de waarde van dit onderzoek is aangetoond zou een hogere onderzoeksbetrouwbaarheid kunnen worden verkregen indien het model bij meerdere organisaties zou worden toegepast en op basis van de bevindingen daaruit waar nodig verder wordt verbeterd.

## Summary

In this research an attempt has been made to answer the question whether a maturity model can be developed for SME organizations to support Data Analytics. In finding an answer, first literature was consulted as a first in-depth study in maturity models.

During the literature study it became clear that maturity models are less used within SME organizations. The lack of specific knowledge and limited resources is cited as the main cause. In addition, the type of maturity model could also be of influence.

There are three main types of maturity models, each with advantages and disadvantages. The Focus Area model, the Staged 5-level model and the Continuous 5-level model. In this study, based on the literature study, the type of Focus Area Maturity model was chosen. This model requires the least knowledge and does not require unnecessary obligations to be met. These features make the model more accessible to SMEs, not to say that it would make it less suitable for larger organizations.

The developed Focus Area Maturity model, based on the Design Science Research model, focuses on Data Analytics. Literature indicates that data availability and data quality are important data characteristics. With these characteristics as focus areas in the model, several features have been described per focus area. Subsequently, points for improvement are described that support the wanted increase of the maturity of a characteristic.

The model has been validated by a single case study. This shows that the model is suitable for application within an SME organization. Consultants can use the model to support their advice towards an SME organization. Use of the model by SME organizations themselves is limited and requires knowledge.

In addition to determining data maturity in SME organizations, the result of this research contributes to the general discussion within Information System (IS) groups about maturity models in general.

Although the added value of this research has been demonstrated through the case study, a higher research reliability will be obtained if the model is applied to multiple organizations and improved where necessary.

# Inhoudsopgave

Abstract .....	ii
Sleutelbegrippen .....	ii
Samenvatting .....	iii
Summary .....	iv
Inhoudsopgave .....	v
<b>1. Inleiding.....</b>	<b>1</b>
1.1. Aanleiding .....	1
1.2. Probleemstelling .....	1
1.3. Onderzoeksvraag .....	2
1.4. Relevantie .....	2
1.5. Onderzoeksaanpak in hoofdlijnen .....	3
<b>2. Theoretisch kader .....</b>	<b>5</b>
2.1. Onderzoeksaanpak.....	5
2.2. Uitvoering.....	6
2.3. Resultaten en conclusies.....	7
2.3.1. Gegevensbeschikbaarheid en gegevenskwaliteit .....	7
2.3.2. Maturity modellen en Data Analytics voor het MKB .....	12
2.4. Doel van het vervolgonderzoek .....	13
<b>3. Methodologie.....</b>	<b>15</b>
3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n) .....	15
3.2. Technisch ontwerp: uitwerking van de methode .....	16
3.3. Rigor en Relevance.....	17
<b>4. Ontwerp, demonstratie, evaluatie en resultaten .....</b>	<b>19</b>
4.1. Iteratie 1.....	19
4.1.1. Ontwerp .....	19
4.1.2. Demonstratie .....	22
4.1.3. Evaluatie.....	23
4.2. Iteratie 2.....	23
4.2.1. Ontwerp .....	23
4.2.2. Demonstratie .....	24
4.2.3. Evaluatie.....	24
4.3. Iteratie 3.....	25

4.3.1.	Ontwerp .....	26
4.3.2.	Demonstratie .....	27
4.3.3.	Evaluatie.....	28
4.4.	Algehele evaluatie.....	28
5.	Discussie, conclusies en aanbevelingen.....	29
5.1.	Discussie.....	29
5.2.	Conclusies .....	30
5.3.	Aanbevelingen voor de praktijk .....	31
5.4.	Aanbevelingen voor verder onderzoek.....	31
	Referenties .....	33
	Bijlage 1, Methodologie van Peffers et al., (2007) DSRModel .....	36
	Bijlage 2, Algemene beschrijving Okoli Model.....	37
	Bijlage 3, Literatuur analyse.....	38
	Bijlage 4, Resultaat na extraction fase (Data mining resultaat onder groene titels).....	39
	Bijlage 5, Libes 2015, databeschikbaarheidsbelemeringen .....	40
	Bijlage 6, Merino 2015, 3 A's Data Kwaliteit-in-gebruik model .....	41
	Bijlage 7, Kenmerken MKB organisatie, Kartiwi et al., 2007.....	42
	Bijlage 8, MKB kenmerken in relatie tot type maturity model .....	44
	Bijlage 9, De zeven richtlijnen van Hevner (2004), p. 12 .....	45
	Bijlage 10, Interview vragen uit de 2e iteratie.....	46
	Bijlage 11, Artefact Dimensie keuzes.....	47
	Bijlage 12, 1 <sup>e</sup> iteratie, benaderde experts en feedback.....	53
	Bijlage 13, Het FAM model, bij 1 <sup>e</sup> iteratie toegepast.....	55
	Bijlage 14, Transcriptie documenten van drie interviews .....	76
	Bijlage 15, 2e iteratie, interview medewerker MKB .....	77
	Bijlage 16, Overzicht van feedback en aanpassingen uit tweede iteratie. ....	78
	Bijlage 17, Het FAM model na twee iteraties. ....	81
	Bijlage 18, MKB kenmerken (Kartiwi & MacGregor (2007) i.c.m. dimensies. ....	102

## 1. Inleiding

In hoofdstuk 1 het onderwerp van dit onderzoek nader beschreven. Als eerste beschreven de aanleiding, gevolgd door de probleemstelling en onderzoeksvraag. Aansluitend de relevantie en in hoofdlijnen de aanpak van het onderzoek.

### 1.1. Aanleiding

Bedrijven verzamelen bij het uitoefenen van hun bedrijfsactiviteiten in de loop der jaren steeds meer data en daarmee neemt de omvang van de verzameling met gegevens ook toe. Deze alsmear groeiende hoeveelheid data, Big Data genoemd, levert op zichzelf weinig informatie voor organisaties op. Analyse van deze grote hoeveelheden gegevensdata daarentegen, kan resulteren in het ontdekken van verbanden, het herkennen van patronen en het opdoen van nieuwe inzichten. Dat kan waardevolle informatie voor een organisatie opleveren. Deze vorm van analyse, ook wel Data Analytics genoemd (hierna: DA), houdt in dat Big Data getransformeerd wordt naar beslissing ondersteunende informatie (Comuzzi & Patel, 2016). Veel organisaties gebruiken tegenwoordig DA om te kunnen (blijven) concurreren (Berndtsson, Forsberg, Stein & Svahn, 2018).

DA-systemen leveren bedrijven voordelen op, die verband houden met bedrijfsprocessen, organisatie performance en concurrentie positie (Davenport, Harris, De Long & Jacobson, 2001). DA kan ingedeeld worden in drie categorieën analyse: descriptive, predictive en prescriptive analyses (Berndtsson et al., 2018). Descriptive analyses helpen organisaties om te analyseren wat er in het verleden is gebeurd. Met predictive analyses kunnen organisaties tot dan toe onbekende patronen in hun gegevens ontdekken met behulp van datamining tools. Prescriptive analyses helpen organisaties bij het automatiseren van te nemen beslissingen, daarbij gebruik makend van eerder opgedane zakelijke inzichten (Berndtsson et al., 2018). Het analyseren van data voor descriptive, predictive en prescriptive toepassingen ter ondersteuning van besluitvorming is steeds succesvoller aan het worden in de praktijk en heeft geleid tot aanzienlijke prestatieverbeteringen bij diverse grote organisaties (Parra, Tort-Martorell, Ruiz-Viñals & Álvarez-Gómez, 2019).

Er is in onderzoeksliteratuur veel informatie beschikbaar die gericht is op Big Data, maar er is veel minder kennis over hoe DA-mogelijkheden kunnen worden toegepast in organisaties (Gupta & George, 2016; Arunachalam, Kumar & Kawalek, 2018). De beperkte kennis over hoe DA toepasbaar is bij besluitvorming weerhoudt organisaties er echter niet van om deze toch te gebruiken. In de literatuur is een toename van het gebruik van DA gesignaleerd (Mikalef, Pappas, Krogstie & Giannakos, 2018) en het zijn vooral de grotere organisaties waar dat gebeurt (Coleman, Göb, Manco, Pievatolo, Tort-Martorell & Reis, 2016). Ook het benutten van de voordelen blijkt binnen MKB-organisaties lastig (Bianchini & Michalkova, 2019).

### 1.2. Probleemstelling

Binnen het MKB blijken organisaties moeite te hebben met het toepassen van DA. Deze bedrijven slagen er veelal niet in om DA succesvol te implementeren en om een deel van het potentieel daarvan te benutten in hun onderneming (Ogbuokiri, Udanor & Agu, 2015). Daarbij is het volgende van belang.

Het lijkt erop dat MKB-organisaties moeite hebben om te voldoen aan de eisen die een succesvolle toepassing van DA stelt aan de verzamelde gegevens. De eisen welke aan verzamelde gegevens



gesteld moeten worden zijn gegevensbeschikbaarheid en gegevenskwaliteit (Melo & Machado, 2019). Bij het toepassen van DA is inzicht in het IT-landschap en alle relevante functies een basisvereiste (Melo & Machado, 2019). De kwaliteit van data-analyses kan nooit uitstijgen boven de kwaliteit van het materiaal in de basis, namelijk de ruwe data.

De gegevensbeschikbaarheid is bij veel MKB-organisaties beperkt, en vaak bestaat die enkel uit interne data. Het toepassen van DA op een dergelijke kleine set van databronnen leidt veelal tot een analyse van slechte kwaliteit (Lazarova-Molnar, Mohamed & Al-Jaroodi, 2019). Alleen door voldoende interne en externe gegevens te gebruiken kunnen organisaties hun vermogen verbeteren om trends gedetailleerder waar te nemen (Bianchini et. al., 2019).

Daarnaast is de gegevenskwaliteit een noodzakelijke basiseis. De vraag wanneer data van dermate goede kwaliteit is dat deze toepasbaar is voor DA is niet in zijn algemeenheid te beantwoorden. Het antwoord op die vraag is sterk afhankelijk van de context en verschilt per situatie (Strong, Lee & Wang, 1997). Datakwaliteit wordt in de literatuur vaak door middel van dimensies beschreven. Eisen die aan datakwaliteit gesteld kunnen worden zijn in paragraaf 2.3 verder beschreven.

### 1.3. Onderzoeksvraag

Zoals gezegd zijn gegevensbeschikbaarheid en gegevenskwaliteit twee belangrijke voorwaarden voor het slagen van DA. Uit de literatuur blijkt dat MKB-organisaties moeite hebben met het toepassen van DA. Maturity modellen zijn een hulpmiddel dat organisaties kan ondersteunen bij onder andere een gefaseerde verbetering van diverse processen. Veel van de bestaande maturity modellen zijn ontwikkeld binnen grote organisaties. MKB-organisaties hebben moeite met het toepassen daarvan. Een maturity model dat gericht is op gegevensbeschikbaarheid en gegevenskwaliteit en dat toepasbaar is binnen het MKB kan mogelijk een oplossing bieden voor het probleem van het achterblijven van succesvol gebruik van DA binnen die categorie ondernemingen .

**De hoofdvraag:** Hoe kan een maturity model, gericht op gegevensbeschikbaarheid en gegevenskwaliteit, ontwikkeld worden ter ondersteuning van Data Analytics en dat bovendien toepasbaar is binnen MKB-organisaties?

Om antwoord te krijgen op bovenstaande hoofdvraag zijn de volgende onderzoeksvragen opgesteld:

**Deelvraag:** Welke eisen kunnen aan gegevensbeschikbaarheid en gegevenskwaliteit worden gesteld voor een MKB-organisatie bij toepassing van Data Analytics?

**Deelvraag:** Welke typen maturity modellen zijn er en waar dient een maturity model gericht op Data Analytics aan te voldoen om bruikbaar te zijn binnen een MKB-organisatie?

### 1.4. Relevantie

Veel bestaande literatuur over maturity modellen is gericht op en ontwikkeld met het oog op toepassing binnen grote organisaties. In het kader van dit onderzoek is geen literatuur gevonden waarin de kenmerken van gegevensbeschikbaarheid of gegevenskwaliteit in een maturity model worden beschreven. Wel is er literatuur waarin gegevensbeschikbaarheid of gegevenskwaliteit als belangrijke factorgenoemd wordt binnen data management modellen, maar daarbij wordt niet verder ingegaan op kenmerken van die factoren zelf (Comuzzi & Patel, 2016; Hardwick et al., 2018;

Immonen et al., 2015; Lui et al., 2020; Woodall & Wainman, 2015). Het DAMA project (Data Management Association) is een project dat weliswaar gericht is op de kwaliteit van informatie, maar het laat de kwaliteit van data (gegevens) helaas, buiten beschouwing. Dat project richt zich voornamelijk op organisaties die informatie leveren, verwerken en afnemen (DAMA-NL, 2019). Ook is er literatuur over data management proces kwaliteit. Het betreffende onderzoek richtte zich op gegevenskwaliteit in de architectuur. Ook in dat onderzoek kwamen specifieke kenmerken van gegevenskwaliteit niet aan bod (Ryu, Park & Park, 2006).

Veel bestaande maturity modellen in de geraadpleegde literatuur kennen een vijftal fasen. MKB-organisaties bereiken in dat stappenschema zelden een hogere fase. Deze modellen blijken binnen het MKB erg moeilijk te implementeren. Ze werken ontmoedigend omdat een volgend niveau voorwaarden stelt die niet van toepassing zijn op de doorsnee MKB-organisatie. Bedrijven haken dan af omdat de gestelde voorwaarden voor hun bedrijf of bedrijfssector niet relevant worden geacht (Mijnhardt, Baars & Spruit (2016). Het toepassen van deze grotendeels complexe maturity modellen vergt zowel tijd als kennis. MKB-organisaties hebben door hun relatief kleine personeelsbestand slechts beperkt de beschikking over tijd, kennis en middelen om de daarvoor benodigde capaciteit in te huren.

Dit onderzoek is relevant vanuit meerdere invalshoeken. In de eerste plaats draagt het bij aan de ontwikkeling van een maturity model dat geschikt is voor MKB-organisaties. Het is gericht op gegevensbeschikbaarheid en gegevenskwaliteit ter ondersteuning van DA. Daarnaast draagt het bij aan het vergroten van de nu nog beperkte wetenschappelijke kennis op het gebied van maturity modellen die specifiek gericht zijn op de, bij de toepassing van die modellen, achterblijvende MKB-organisaties. Het onderzoek vergroot en verdiept het inzicht in bestaande maturity modellen en de geschiktheid daarvan voor het MKB. Het draagt verder bij aan de kennis op het terrein van data management, in het bijzonder op het verzorgen van aanwezige data. Goed verzorgde data zijn compleet, betrouwbaar en tijdig en daardoor geschikt voor toepassingen die van die data gebruik maken.

Naast de relevantie van dit onderzoek voor de al genoemde deelonderwerpen binnen de wetenschap heeft het ook relevantie voor de praktijk. Het levert op verschillende manieren een bijdrage aan het functioneren van MKB-organisaties. Het ondersteunt om te beginnen bij de beschrijvende (descriptive) analyse. De maturity meting (As-Is situatie bepalen) vindt plaats door middel van stellingen en is met beperkte kennis van dit onderwerp betrekkelijk makkelijk toe te passen. Het model levert, na de beschrijvende fase, de mogelijkheid om eenvoudig keuzes te maken op kenmerkenniveau. Dit onderzoek helpt op de vlakken waar verbetering in de bedrijven gewenst wordt geacht en reikt aan welke doelgerichte acties vervolgens kunnen worden ondernomen. Hoe beter een MKB-organisatie in staat is om de data op een goede manier te gebruiken, des te groter zijn de voordelen die behaald kunnen worden. Het resultaat kan zich vertalen in lagere kosten, hogere winsten en concurrentievoordelen. Essentieel daarvoor is dat een maturity model begrijpelijk en toepasbaar is.

## 1.5. Onderzoeksaanpak in hoofdlijnen

Om tot een toepasbaar maturity model (het artefact) te komen volgt dit onderzoek de design science in research methodologie (Peffer et al., 2017). Design science houdt in het creëren en evalueren van IT-artefacten en is er op gericht om problemen binnen organisaties op te lossen. Het design science research model (hierna: DSR-model) van Peffer et al., (2017) kent een zestal

stappen: 1) probleemidentificatie en motivatie; 2) doelstelling; 3) ontwerp en ontwikkel; 4) demonstratie in de vorm van testen; 5) evaluatie en 6) communicatie.  
Voor een gedetailleerde beschrijving zie hoofdstuk 3.2 en bijlage 1.

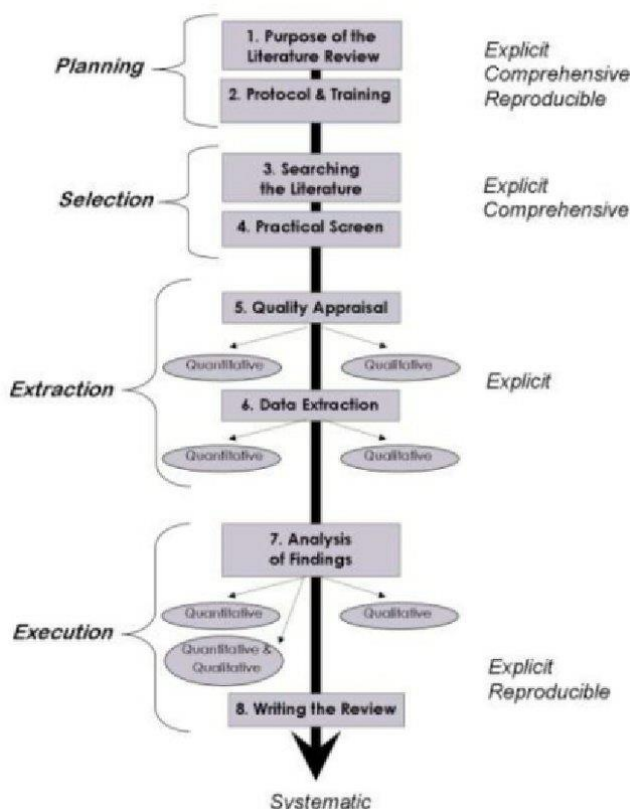
## 2. Theoretisch kader

In dit hoofdstuk, het theoretisch kader, wordt de belangrijke bestaande theoretische kennis kort beschreven. Hiermee wordt inzicht verkregen in bestaande kennis op het gebied van gegevensbeschikbaarheid, gegevenskwaliteit en het ontwikkelen en toepassen van een DA maturity model voor MKB-organisaties.

### 2.1. Onderzoeksaanpak

Om onderzoeksvraag (met inbegrip van de deelvragen) te kunnen beantwoorden zal eerst een zogenaamde literatuurreview uitgevoerd worden. Met behulp daarvan wordt inzichtelijk gemaakt welke typen maturity modellen er zijn, welke kenmerken bevatten deze typen maturity modellen en welke type leent zich het meest geschikt voor toepassing binnen MKB-organisaties. Daarnaast wordt helder gemaakt welke voorwaarden gesteld kunnen worden aan gegevensbeschikbaarheid en gegevenskwaliteit voor MKB-organisaties in relatie tot DA. Uiteindelijk zal op basis van de uitkomsten van het literatuuronderzoek een maturity model ontwikkeld worden en een toetsing in een praktijksituatie om de kwaliteit te van het ontwikkelde model verder te verbeteren.

Bij dit onderzoek zal gebruik worden gemaakt van de methode van de structurele literatuuranalyse (SLR) zoals beschreven door Okoli (2015), zie figuur 2. Naast het SLR model van Okoli bestaat er een tweede systematisch review, de structurele mapping studie (SMS) (Napoleão, Felizardo, De Souza, & Vijaykumar, 2017). Vergelijking van de beide methodes maakt duidelijk dat kwaliteitsbeoordeling breder is toegepast in SLR en om die reden valt de keuze binnen dit onderzoek op de SLR-methode. Zoals in de figuur is weergegeven, bestaat deze methode uit een achttal stappen. Op een hoger niveau kunnen deze acht stappen worden gegroepeerd in een viertal opeenvolgende fasen; In paragraaf 2.2 worden de fasen Planning (*Planning*), Selectie (*Selection*) en Samenvatten (*Extraction*) beschreven. Uitvoeren (*Execution*) wordt in paragraaf 2.3 nader beschreven.



## 2.2. Uitvoering

Zoals in de vorige paragraaf is gemeld, volgt dit onderzoek het Okoli model zoals weergegeven in Figuur 2. Voor een algemene beschrijving van het model zie bijlage 2.

### Fase 1 Planning

De eerste stap van de planning bestaat uit twee onderdelen, te weten: de beschrijving van het doel van het literatuuronderzoek en het opstellen van een protocol en een training.

Doel van het literatuuronderzoek is het verkrijgen van inzicht in reeds bestaande literatuur, specifiek gericht op het gebied van de begrippen in de query. Het protocol en de training zijn gericht op de toepassen binnen onderzoeksteams. Het zorgt voor gelijke criteria tussen leden van het onderzoeksteam tijdens de uitvoering van het literatuuronderzoek. Aangezien bij dit onderzoek geen sprake is van meerdere onderzoekers is dit hier niet van toepassing en behoeft het geen verdere uitwerking.

### Fase 2 Selectie

Ook de stap van de selectie kent twee onderdelen. Eerst wordt gezocht naar relevante literatuur en daarna volgt de zogenaamde Practical Screen, met andere woorden de inhoudelijke beoordeling van de gevonden literatuur.

Het zoeken naar *literatuur is gedaan door* gebruik te maken van de database van de Open Universiteit (hierna: OU). Daarbij is binnen de zoekresultaten omwille van de kwaliteit gefilterd op:

- Peer reviewed publication.
- Content type: Dissertation/Thesis, Journal Article, Paper en Publication.
- Vinkje 'tonen op relevantie' bovenaan op de resultaten pagina is geselecteerd.
- Toevoegen van resultaten buiten de bibliotheek collectie.

Om mogelijke publicaties buiten de OU database(s) toch te kunnen vinden is dit laatst genoemde criterium optie geselecteerd. De toegepaste query is in Tabel 1 weergegeven.

Uit eigen waarneming blijkt dat het aantal publicaties de laatste jaren is gestegen. Daarom is met het oog op relevantie besloten een datum criterium toe te voegen.

- Datum criterium: publicaties van de laatste vijf jaar.

Tabel 1 Query begrippen en de query

In Query begrippen:	Data Analytics, Business Intelligence, Information Availability, Information Quality, Data Availability, Data Quality, SME, Maturity en Mature
Query	<pre>((SME) AND((Data Availability) OR (Data Quality)) OR ((Information Availability) OR (Information Quality))OR ((Data Analytics) OR (Business Intelligence))) OR ((SME) AND ((Maturity) OR (Mature))))</pre> <p>Deze query levert 50.197 gevonden items op.</p>

Vervolgens is gestart met het analyseren van de titels uit de zoekresultaten tot het punt werd bereikt dat er op minimaal 3 pagina's (met telkens à 25 titels) geen relevante titels meer werden aangetroffen.

*De Practical Screen bestond uit*, het beoordelen van de literatuur op de lijst van zoekresultaten met relevante titels. Beoordeeld moest worden, welke literatuur daaruit gefilterd moest worden voor verdere analyse en welke literatuur niet zou worden betrokken bij het onderzoek. Tijdens het beoordelen van de literatuur is door middel van de zogenaamde snowballing methode nog extra literatuur aangetroffen. Deze vondsten leveren mogelijk een waardevolle toevoeging aan dit onderzoek. In eerste instantie zijn deze via de OU-database gezocht en voor zover de zoekopdracht daar geen resultaat op leverde is via Google Scholar nogmaals een zoekslag gedaan. Zie bijlage 3 de literatuurlijst met de relevante titels en vindplaatsen.

### Fase 3 Extraction

Extraction is een cruciale fase binnen de SLR-methode. Deze fase levert de onderzoeker uit de geselecteerde literatuur, op systematische wijze de basis op voor verdere analyse.

Het Okoli model beschrijft deze fase in twee stappen, namelijk Quality Appraisal en Data Extraction. De stap van Quality Appraisal houdt een verdere beoordeling in van de geselecteerde literatuur na de fase van selectie. Het gaat hierbij om een beoordeling van de literatuur op kwaliteit, zoals bijvoorbeeld de betrouwbaarheid van het onderzoek.

De stap van Data Extraction beschrijft data mining van de geselecteerde literatuur zoals het weergeven via welke database het is gevonden, het jaar van publicatie, het publicatie type, zie bijlage 4 voor het resultaat na de extraction fase.

Resultaat is een tabel van 74 (63 +11) items, zie bijlage 4 voor meer detail informatie. In tabel 2 het extractie resultaat kort weergegeven.

*Tabel 2 Extractie resultaat*

Na abstract lezen niet als voldoende beoordeeld:	11
Resterend aantal items waar de inleiding is beoordeeld en waar nodig ook de conclusie is beoordeeld:	63
Resterend uiteindelijk aantal geselecteerd (na extraction):	20

## 2.3. Resultaten en conclusies

Na het doornemen van de geselecteerde literatuur zal in deze paragraaf op basis daarvan zoveel als mogelijk een antwoord gegeven worden op elk van de gestelde onderzoeksvragen. Per vraag zullen telkens de stappen 7 en 8 uit figuur 2 worden beschreven.

### 2.3.1. Gegevensbeschikbaarheid en gegevenskwaliteit

**Deelvraag:** Welke eisen kunnen aan gegevensbeschikbaarheid en gegevenskwaliteit worden gesteld voor een MKB-organisatie bij toepassing van Data Analytics?

#### Definities

Om deze vraag te kunnen beantwoorden is allereerst van belang om helder te definiëren wat in dit onderzoek en in literatuur wordt verstaan onder de begrippen gegevensbeschikbaarheid en gegevenskwaliteit.

Gegevensbeschikbaarheid heeft betrekking op de tijdigheid en de betrouwbaarheid van toegang tot en gebruik van gegevens. In de literatuur wordt gegevensbeschikbaarheid in twee verschillende

contexten beschreven, namelijk 1) de mate waarin data continue beschikbaar is (Immonen et al., 2015), (dit kan vooral met technische middelen worden bereikt) en 2) de mate waarin men over verschillende datasets kan beschikken (Hardwick, 2018; Comuzzi & Patel, 2016). De eerstgenoemde interpretatie valt buiten het bereik van dit onderzoek en blijft daarom hier buiten beschouwing. Gezien vanuit de tweede context wordt databeschikbaarheid gedefinieerd als *de graad van gemak voor gebruikers om gegevens en gerelateerde informatie te verkrijgen* (Costa, Freitas, Stefanik, Krafft, Pilot, Morrison & Santana, 2019). In dit onderzoek wordt deze tweede definitie gehanteerd, omdat deze definitie de moeilijkheidsgraad bevat die gebruikers kunnen ondervinden bij het ontsluiten van verzamelde gegevens. Bovendien staat deze definitie uit 2019 veel dichterbij de actualiteit.

Van gegevenskwaliteit zijn vele definities in literatuur te vinden, die allemaal vergelijkbaar zijn. Haug & Arlbjorn (2010) verwijzen naar een viertal onderzoeken met een klassieke definitie van datakwaliteit 'geschiktheid voor gebruik', dat wil zeggen de mate waarin sommige gegevens met succes de doeleinden van de gebruiker dienen. Dergelijke definities impliceren dat datakwaliteit contextueel is. Een aantal datakwaliteitskenmerken waaronder geloofwaardigheid zijn contextueel. Zie tabel 5, Data Quality-in-Use model for Data Analytics, Merino et al., (2015) voor meer contextuele kenmerken. Een andere, nogal summier, definitie van datakwaliteit is 'gegevens die geschikt zijn voor gebruik door data verwerkers' (Baars, 2016). Meer uitgebreid is 'de mate waarin de kwaliteit voldoet aan, of deze de kwaliteitseisen overtreft, die gebruikers ervan verwachten' (Sandhu, 1993). Dit onderzoek kiest voor de laatstgenoemde definitie omdat deze ruimer is en bovendien de mate van gebruikersbeleving omvat.

Bij de bespreking van beide definities (gegevensbeschikbaarheid en gegevenskwaliteit) zal het begrip gebruikers wat breder beschouwd worden. In de context van DA gaat het niet alleen om dat gebruikers maar ook om de vraag welke eisen de data-analyse (tools) zelf als gebruiker stelt of als benodigde handeling vereist om data te kunnen verwerken.

### Eisen gegevensbeschikbaarheid

Bij het stellen van eisen aan gegevensbeschikbaarheid zijn een aantal factoren van belang, zoals de vraag welke gegevens worden verzameld, de vraag hoe de in snelheid toenemende data worden verzameld en ten slotte de vraag hoeveel data worden verzameld. De antwoorden op deze belangrijke vragen bepalen sterk de gegevensbeschikbaarheid (Coleman et al., 2016; Hardwicke et al., 2018; Immonen et al., 2015; Juddoo, 2015; Libes, 2015). Ook de vraag naar gegarandeerde betrouwbare data-analyse en de factor complexiteit spelen een belangrijke rol bij gegevensbeschikbaarheid. In de context van DA is bovendien van belang of er toegang is tot (externe) data en hoe data tot waardevolle bedrijfsinput wordt getransformeerd (herbruikbaarheid, domein specifiek en begrijpelijk).

Diverse literatuur benoemt databeschikbaarheid als een belangrijke vereiste bij het gebruik van DA (Coleman et al., 2016; Libes, 2015). Nergens in die literatuur wordt echter beschreven welke specifieke eisen daaraan gesteld kunnen worden. Alleen Libes (2015) heeft beschreven welke belemmeringen bestaan bij gegevensbeschikbaarheid in productiebedrijven en heeft een achttal kenmerken onderscheiden, zie Tabel 3, op basis waarvan mogelijk eisen gesteld kunnen worden.

Tabel 3 Kenmerken van gegevensbeschikbaarheid (Libes, 2015)

1 Beperking in gegevensgebruik;	A Bedrijfseigen gegevens, niet toegankelijk voor derden (en onderzoekers). Daarmee is de bijdrage van onderzoekers tot het leveren van een toepasbare oplossing beperkt.
---------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	B Gegevens beveiliging. Ter bescherming van bedrijfsdata, wordt ook binnen de MKB bedrijfsmuren segmentatie van toegang tot data ingericht. Kans op te beperkte data waarop vervolgens analyse wordt toegepast en dit verkeerde resultaten oplevert.
2 Niet bestaande of ontbrekende gegevens	Gegevens kunnen niet verzameld worden vanwege de kosten die dit met zich meebrengt, tijdelijke storing bij data verzamel opslag of storing bij de sensoren.
3 Timing en synchronisatie van gegevens	Bij koppeling van data uit meerdere informatiebronnen is het begrip van wanneer de informatie is gegenereerd belangrijk. De juiste definitie van de vastgelegde tijd is belangrijk om te kunnen vergelijken en koppelen met andere data.
4 Gegevensfrequentie	De frequentie waarmee gegevens worden verzameld. Een lage frequentie levert minder data op en kan leiden tot minder inzichten. Een hogere frequentie vraagt meer opslag ruimte, kan tot performanceproblemen leiden en tot hogere kosten.
5 Gegevensformat, standaarden en specificaties	Organisaties kunnen een uitgebreide verzameling van uiteenlopende gegevens bezitten. Vaak opgeslagen in bestanden die geformatteerd zijn volgens verschillende standaarden met een verschil in de mate van nauwkeurigheid.
6 Onzekerheid, betrouwbaarheid en nauwkeurigheid	Bedrijfsgegevens kunnen allerlei kenmerken hebben bij aanvang van data-analyse. Metadata (extra informatie die de data beschrijven) kan daarbij ondersteunen.
7 Toegang tot gegevens, opslag en verwerking	Gegevenstoegang en beschikbaarheid zijn van invloed op data-analyse. Zelfs in eenvoudige gevallen worden gegevens structureel gewijzigd wanneer ze uitgedeeld worden aan onderzoekers. Zo is de distributie van digitale gegevens traditioneel een serialisatie (omvormen zodat het transporteerbaar of opslaanbaar is).
8 Voorbereidingstijd voor gegevens en andere bronnen	Data verzamelen betekent ook opslag(ruimte) gereed maken voor het bewaren van de data. Welke data waar precies opslaan vergt voorbereiding.

Naast deze kenmerken onderscheidt Libes (2015) ook gegevensbeschikbaarheid naar datacategorieën (datadomeinen) en soorten data. Voorbeeld van datacategorieën volgens Libes (2015) zijn: Inkoop Management Data, Productieplanningsgegevens, Procesplanningsgegevens, Machinedata. Een dergelijke indeling past binnen de aanwezige behoefte aan modellen die meer houvast bieden bij het stapsgewijs verbeteren van een domein (Steenbergen, Bos, Brinkkemper, Van de Weerd & Bekkers, 2010).

Bij de indeling naar soorten data beschrijft Libes (2015) de volgende soorten die elk eigen kenmerken hebben: Key Performance Indicatoren, menselijke interacties, data modellen, hoge kwaliteit data, metadata en gesimuleerde data.

Samengevat kan worden gesteld dat de bestaande literatuur geen specifieke eisen beschrijft voor gegevensbeschikbaarheid. In dit onderzoek zal voor gegevensbeschikbaarheid de beschreven acht-



tal kenmerken (Libes, 2015), zie Tabel 3 en uitgewerkt in bijlage 5, in een maturity model worden verwerkt.

### Eisen gegevenskwaliteit

De eisen die gesteld kunnen worden aan gegevenskwaliteit bij het gebruik van DA zijn sterk context afhankelijk (Juddoo, 2015). Er zijn zoveel manieren waarop data gebruikt en toegepast wordt dat het uitdrukken van datakwaliteit in veel literatuur aan de hand van dimensies gebeurt (Merino, Caballero, Rivas, Serrano & Piattini, 2016). Een klassieke indeling van data kwaliteitskenmerken bestaat uit vier categorieën; intrinsiek, toegankelijkheid, contextueel en representatief. Intrinsieke kwaliteit houdt in dat gegevens op zichzelf kwaliteit hebben, onafhankelijk van de gebruikerscontext. Toegankelijkheid betreft de aspecten die een relatie hebben met de toegang tot informatie. Contextueel beschouwt datakwaliteit binnen de context van toepassing en bevat subjectieve voorkeuren van de gebruiker. Representatieve kenmerken leggen de aspecten vast die verband houden met de informatie vertegenwoordiging (Strong, 1997; Immonen, 2015). De vier categorieën en de daarbij behorende dimensies zijn weergegeven in Tabel 4 Data kwaliteit Categorieën en bijbehorende Dimensies (Strong, 1997).

*Tabel 4 Data kwaliteit Categorieën en bijbehorende Dimensies (Strong, 1997)*

<b>DQ Category</b>	<b>DQ Dimensions</b>
Intrinsic DQ	Accuracy, Objectivity, Believability, Reputation
Accessibility DQ	Accessibility, Access security
Contextual DQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational DQ	Interpretability, Ease of understanding, Concise representation, Consistent representation

Op basis van de klassieke indeling van Strong (1997) is het 'Data Kwaliteit-in-gebruik' model voor DA beschreven (Merino, Ismael, Bibiano, Manuel & Mario, 2015). Daarbij zijn de vier categorieën uit het onderzoek van Strong (1997): toegankelijkheid, contextueel, representatief en intrinsiek, door Merino et al., (2015) ingedeeld in uiteindelijk drie kenmerken: contextuele geschiktheid, operationele geschiktheid en tijdelijke geschiktheid (zie Tabel 5). Deze indeling is specifiek gericht op categoriseren van data voor DA.

*Tabel 5 Data Quality-in-Use model for Data Analytics, Merino et al., (2015)*

<b>Data Quality Characteristic</b>	<b>Contextual Adequacy</b>	<b>Temporal Adequacy</b>	<b>Operational Adequacy</b>
Accuracy	X	X	
Completeness	X		
Consistency	X	X	
Credibility	X		
Currentness		X	
Accessability			X
Compliance	X		
Confidentiality	X		X
Efficiency			X
Precision			X

Traceability			X
Understandability	X		
Availability			X
Portability			X
Recoverability			X

Contextuele geschiktheid in Tabel 5 verwijst naar “het vermogen van datasets om te worden gebruikt binnen hetzelfde interessegebied van de analyse, onafhankelijk van welk formaat dan ook (bijv. gestructureerd versus ongestructureerd), elke grootte of de snelheid van instroom” (Merino et al., 2015). Tijdelijke geschiktheid verwijst naar “data binnen een passend tijdvak voor de analyse” (Merino et al., 2015), (bijv. vergelijkbare leeftijd of gelijktijdige gegevens). Dat betekent dat de gegevens verwijzen naar een soortgelijke periode. De laatste, operationele geschiktheid, betreft de mate waarin gegevens kunnen worden verwerkt in de beoogde analyse en geschikte methode zonder dat er gegevens (deels) buiten de analyse vallen. Dit betekent dat er voldoende en geschikte middelen zijn om de analyse uit te voeren (bijv. vergelijkbare gegevenstypen). Voor een beschrijving van de Data Kwaliteit kenmerken, zie bijlage 6.

In dit onderzoek stellen de onderzoekers dat de toereikendheid van gegevens de belangrijkste factor is waarmee datakwaliteit kan worden beoordeeld. Toereikendheid wordt beschreven als de staat (inherente waarde) of het vermogen van gegevens (systeem afhankelijke waarde) om te voldoen aan de specifieke kenmerken van de analyse die wordt uitgevoerd.

Verder beschrijft de literatuur algemene kenmerken, trends over gegevensbeschikbaarheid en gegevenskwaliteit. Beschreven wordt dat een goede data-analyse sterk afhankelijk is van de kwantiteit en de kwaliteit van gegevens. Voor beide geldt dat ze niet gemakkelijk beschikbaar zijn binnen het MKB (Lazarova-Molnar et al., 2019). Daarvoor is, het aantal interne gegevensbronnen over het algemeen te beperkt. Wel wordt gesignaleerd dat de bewustwording toeneemt van het belang van (interne) data verzameling (Comuzzi & Patel, 2016) en daarnaast een toename onder invloed van concurrentiedruk (Coleman et al., 2016). Onderzoekers wijzen op het belang van de toegang tot externe gegevensbronnen (English & Hoffmann, 2018). Via open (statistische) informatiebronnen zoals bijvoorbeeld het Centraal Bureau voor de Statistiek (CBS), kunnen bedrijven informatie verkrijgen. Organisaties zijn niet langer slechts toeleverancier van informatie, maar ze krijgen een groeiend aandeel als afnemers van statistische informatie (Bavdaž, Bergström, Biffignandi, Bolko, Giesen, Gravem, Löfgren, Persson, Mohorič Peternelj & Torres van Grinsven, 2011; Coleman et al., 2016). Het verwerken van deze informatie vereist specifieke kennis. Sommige onderzoekers zien een andere mogelijke oplossing in Collaborative Data Analytics (hierna: CDA's) (Lazarova-Molnar et al., 2019). CDA's stellen een MKB in staat om data-analyse processen te vergemakkelijken. CDA is een samenwerking van MKB-organisaties waarbij de gegevens gezamenlijk worden verwerkt en geanalyseerd en resultaten met alle deelnemers worden gedeeld. Dit wordt gezien als een uiterst krachtige manier om nuttige inzichten te verkrijgen (Vecchio et al., 2017). Coleman (2016) verwijst naar een vereniging van openbare gegevens. Voorwaarde voor succes hierbij is een 'open' benadering om de enorme hoeveelheid gegevens real-time te kunnen verwerken. Hierin ligt tegelijkertijd de uitdaging voor MKB's. Voorbeelden hierbij welke het bemoeilijken zijn, AVG-wetgeving, het spanningsveld tussen openheid en concurrentie, juridische aspecten, etc. De grens tussen bedrijven wordt hierbij poreuzer en coöperatiever (Vecchio et al., 2017).

De belangrijkste doelstelling van alle DA systemen is het vereenvoudigen van het besluitvormingsproces door de verstrekking van kwaliteitsinformatie (Papachristodoulou et al., 2017). Onder invloed van technologische ontwikkelingen zijn inmiddels MKB-toepasbare online DA tools in cloud systemen voorhanden (Coleman, 2016). Cloud systemen worden in diverse onderzoeken aanbevolen als (extern) middel voor het MKB om informatie te verwerken tot kwalitatieve data (Coleman, 2016; English et al., 2018; Liu et al., 2020; Papachristodoulou et al., 2017).

Cloudservers bieden MKB-organisaties de mogelijkheid om complexe data te verwerken, wat normaal gesproken met de beperkingen die aan de bestaande middelen kleven niet uitvoerbaar is. Een nadeel dat verbonden blijft aan cloudservers is de benodigde kennis voor de verwerking van data en die rechtstreeks van invloed is op de gegevenskwaliteit.

De kwaliteit van data-analyse kan maximaal gelijk zijn aan de kwaliteit van de beschikbare data. De zwakste schakel bepaalt de kracht van de keten.

Problemen waar MKB-organisaties tegen aanlopen bij data-analyse zijn beslissingen met betrekking tot dataverzameling (de software, de kennis en het format). De hoeveelheid beschikbare data is van grote invloed bij de toepassing van DA. Belangrijke beslissingen betreffen de databronnen, het datatype, de datalocatie en de mate van verzamelen. Het ontbreekt veelal aan intuïtieve DA software (Coleman, 2016). Bestaande DA software is grofweg in twee categorieën onder te brengen:: enerzijds een categorie van goed bruikbare maar complexe software die specifieke kennis vereist en anderzijds de categorie van eenvoudige maar slechts beperkt effectieve software.

### 2.3.2. Maturity modellen en Data Analytics voor het MKB

**Deelvraag:** Welke typen maturity modellen zijn er en waar dient een maturity model gericht op Data Analytics aan te voldoen om bruikbaar te zijn binnen een MKB-organisatie?

#### **Maturity modellen**

Een maturity model onderzoekt in de basis een (deel)proces binnen een organisatie.

Er kan onderscheid gemaakt worden naar een drietal verschillende maturity modellen (Steenbergen, 2011): het Staged 5-level model (fixed level), het Continuous 5-level model (fixed level) en het Focus area model.

Bij het Staged 5-level model moeten de focusgebieden naar tevredenheid worden geïmplementeerd, om te verzekeren dat de organisatie dat maturity niveau daadwerkelijk bereikt. Bij het Continuous 5-level model worden de focusgebieden niet aan een niveau toegeschreven, maar worden binnen het kader van elk afzonderlijk focusgebied vijf niveaus onderscheiden. Beide laatstgenoemde modellen gaan uit van vijf generieke niveaus.

Baars et al., (2016) onderzocht een groot aantal bestaande fixed-level maturity modellen (Staged 5-level model en Continuous 5-level) en onderscheid een drietal type modellen, namelijk; Maturity rasters, Likert-schaal vragenlijst en Capability Maturity Model (CMM)-achtige modellen. In de meeste literatuur over Fixed-level maturity wordt een vijftal niveaus met verschillende dimensies onderscheiden. Fixed-level maturity modellen met een vast niveau zijn zeer geschikt voor benchmarking, waarbij een organisatie op volwassenheidsniveau wordt geplaatst door te beoordelen in hoeverre de bijbehorende processen zijn geïmplementeerd, maar ze zijn helaas minder geschikt voor incrementele verbetering. Ze kunnen namelijk geen onderlinge afhankelijkheid uitdrukken tussen de processen die de volwassenheidsniveaus vormen (Steenbergen et al., 2010). Het Focus area model gaat uit van een eigen aantal volwassenheidsfases per focusgebied. De mate van volwassenheid wordt hierbij beschreven als een combinatie van niveaus binnen deze

focusgebieden. Alle modellen kennen het principe van erven. Wil men aan het volgende niveau kunnen beginnen dan dient men eerst het daaraan voorgaande niveau te hebben bereikt. In dit onderzoek wordt het Focus Area Maturity (FAM) model toegepast. Het model is uitgebreid en gedetailleerd, maar bevat in de basis niet de kenmerken van de organisatie zoals de omvang van het personeelsbestand, de omzet, de geografische locatie en aard van de branche. Geadviseerd wordt om deze kenmerken juist wel mee te nemen (Mijnhardt et al., 2016) omdat bijvoorbeeld op het vlak van risico en bedreigingen belangrijke verschillen bestaan tussen bijvoorbeeld MKB met 2 of 200 personeelsleden.

### **Maturity modellen voor het MKB**

Veel MKB-organisaties zijn over het algemeen niet in staat om een hoger volwassenheidsniveau te bereiken omdat de maturity vereisten voor zo'n volgend niveau te moeilijk zijn te implementeren of te veel investering vergt (op het gebied van tijd, kennis, en financiële middelen). De vereisten blijken een te grote stap om een volgend maturity niveau te behalen. De plicht om te voldoen aan kenmerken of capaciteiten die niet van toepassing zijn op het bedrijf of niet relevant worden geacht binnen de betreffende bedrijfssector werkt ontmoedigend (Mijnhardt et al., 2016). CMM-achtige modellen zijn gezien de complexiteit (Baars et al., 2016) minder geschikt voor het MKB. Een FAM model kenmerkt zich door kleinere tussenstappen tussen de verschillende maturity fasen per focusgebied en lijkt om die reden veel beter geschikt voor toepassing binnen MKB-organisaties. Een dergelijk model biedt meer begeleiding op incrementele verbetering dan fixed level maturity modellen (Steenbergen et al., 2010). Ook niet onbelangrijk is dat ze bedrijven de mogelijkheid bieden om specifieke focusgebieden beperkt toe te passen. Er is dus meer ruimte voor het bieden van maatwerk. Het toepassen van organisatiekenmerken op focusgebieden valt overigens buiten de scope van dit onderzoek. De MKB-organisatie kenmerken (Kartiwi et al., 2007), beschreven in bijlage 7, zijn in bijlage 8 in een tabel opgenomen. Daarin is de relatie tussen MKB kenmerk en de drie typen maturity modellen beschreven.

Een FAM model (Steenbergen et al., 2010) is gebaseerd op het concept van een aantal focusgebieden die ontwikkeld moeten worden om volwassenheid te bereiken in een functioneel domein. Men definieert het focusgebied als een aspect dat tot op zekere hoogte moet worden geïmplementeerd om een functioneel domein effectief te laten zijn. Maturity modellen gebaseerd op FAM bieden meer begeleiding voor incrementele verbetering ten opzichte van modellen gebaseerd op Fixed-level (Steenbergen, 2011). Een bestaand FAM model is het ISFAM (Information Security Area Maturity) model (Mijnhardt et al., 2016) en is specifiek gericht op MKB-organisaties.

FAM modellen lijken geschikt voor toepassing bij MKB organisaties. Uit het literatuuronderzoek is gebleken dat het daarin tot nu toe ontbreekt aan gevalideerde FAM-modellen gericht op gegevensbeschikbaarheid en op gegevenskwaliteit voor succesvolle toepassing binnen het MKB.

## **2.4. Doel van het vervolgonderzoek**

Het doel van dit onderzoek is het ontwerpen van een gevalideerd maturity model om DA binnen het MKB te ondersteunen. Het zwaartepunt van het model zal liggen bij gegevensbeschikbaarheid en gegevenskwaliteit. De beschikbaarheid en kwaliteit van gegevens vormen de noodzakelijke basis voor het toepassen van DA. Bestaande literatuur wijst uit dat veel bestaande maturity modellen ongeschikt zijn voor het MKB (Coleman et al., 2016; Lazarova-Molnar, 2019; Mittal, 2018). In het MKB worden over het algemeen slechts in beperkte mate (interne data) gegevens vastgelegd en

bovendien slechts beperkt gebruik gemaakt van externe data (Lazarova-Molnar et al., 2019). Het ontbreekt in de bestaande literatuur aan relevante richtlijnen voor het verzamelen van gegevens, voor de benodigde soorten gegevens en voor de hoeveelheid gegevens (Libes et al., 2015) om met succes gebruik te kunnen maken van DA.

In hoofdstuk 3 wordt beschreven hoe een model is ontwikkeld met gebruik van het DSR-model (Peffers et al., 2007) met drie iteraties waarvan de laatste via een case study.

### 3. Methodologie

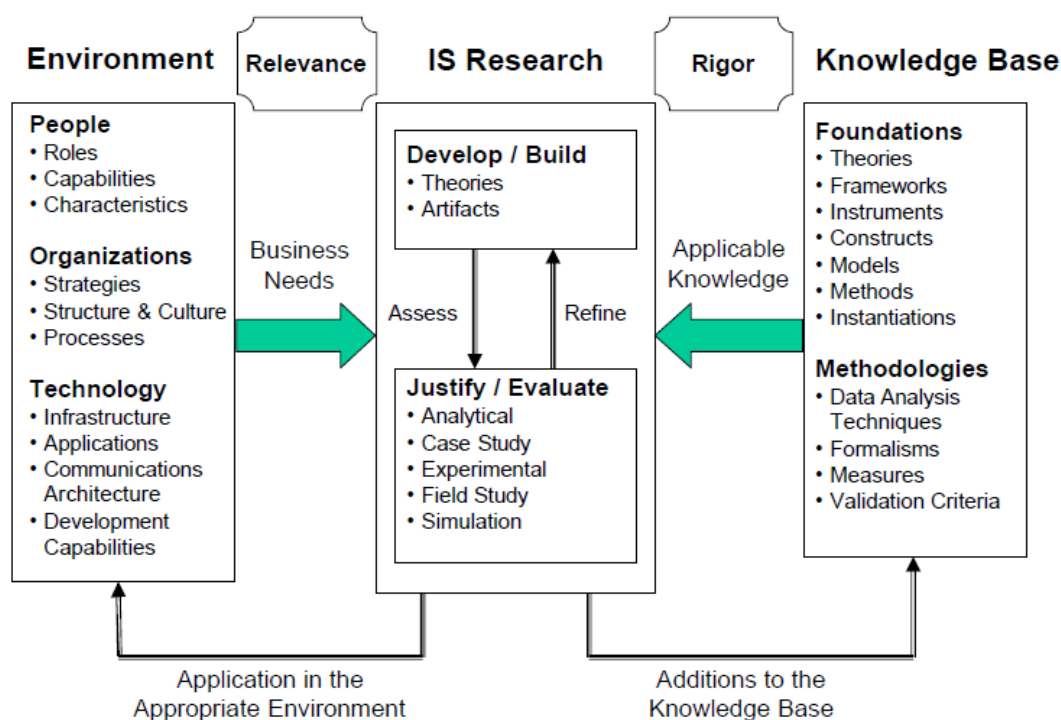
Door gebruik te maken van empirische onderzoek zal de onderzoeksvraag op verantwoorde wijze worden beantwoord. De toegepaste onderzoeksmethode is in dit hoofdstuk nader beschreven.

#### 3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Dit onderzoek richt zich op het ontwikkelen van een maturity model, specifiek voor het MKB ter ondersteuning van DA.

De basis voor dit onderzoek is het DSR-model (Hevner, 2004), zie figuur 3. Dat geeft in essentie een raamwerk voor het begrijpen, uitvoeren en evalueren van IS onderzoek en de samenhang tussen gedrags- en ontwerpwetenschappen. Het model kort beschreven: Development & Justify, als gedragswetenschap voor het vinden van de waarheid, gekenmerkt als reactief. Build & Evaluate als ontwerpwetenschap, gekenmerkt als proactief. Samen binnen IS onderzoek om via Assess een artefact te beoordelen en een verscherping via Refine.

Het uiteindelijke artefact zal hierbij beoordeeld worden op het toevoegen van kennis aan de wetenschap en het probleem oplossen aan de omgeving (het MKB).



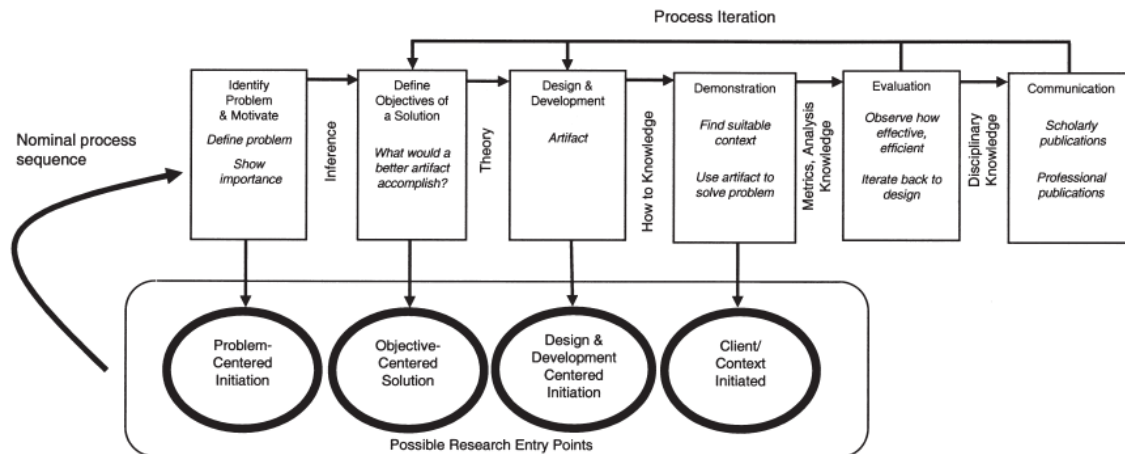
Figuur 3 Schematisch DSR-model. Kopie uit Hevner et al., (2004)

Dit onderzoek volgt het DSR-model (Peffer et al., 2007), zie Figuur 4. In bijlage 1 staat het DSR-model met extra beschrijving. Het DSR-model richt zich op wetenschappelijk onderzoek en levert een artefact op, in dit onderzoek het maturity model. Een relevante oplossing voor een bestaand probleem, namelijk het niet voorhanden zijn van een geschikt maturity model voor het MKB. Via deze methodologie zal het gedemonstreerd en geëvalueerd worden en daarmee bijgedragen worden aan de validiteit. Deze methode van onderzoek is goed toepasbaar in dit onderzoek. Het evalueren van het artefact is een kernactiviteit in DSR als basis voor verdere ontwikkeling en draagt bij aan de nauwkeurigheid van het artefact (Venable et al., 2016).

Uiteindelijk zal een case study een toetsing in de praktijk vormen en als basis dienen voor de evaluatie in hoofdstuk 4. Ten slotte moeten de resultaten gecommuniceerd worden.

### 3.2. Technisch ontwerp: uitwerking van de methode

Het DSR-model (Peffer et al., 2007), zie figuur 4 nader beschreven (grotere weergave in bijlage 1). Dit model bevat een zestal stappen welke structuur bieden bij het ontwikkelen en evalueren van een artefact. Hieronder het model met aansluitend een korte beschrijving van de stappen en vervolgens een aanvulling op elke stap met concrete onderzoeksstappen binnen het kader van dit onderzoek.



Figuur 4: DSR-model, Peffer et al., (2007)

#### Stap 1, *Probleemidentificatie en motivatie.*

Deze stap, omvat het definiëren van het specifieke onderzoeksprobleem en de motivatie of rechtvaardiging van een oplossing daarvoor.

Literatuuronderzoek in hoofdstuk 2 beschrijft het gemis van een maturity model voor het MKB gericht op gegevensbeschikbaarheid en gegevenskwaliteit.

#### Stap 2, Het afleiden van de *doelstellingen*.

Het betreft hier de doelstellingen (zowel kwantitatief als kwalitatief) voor een oplossing uit het onderzoeksprobleem. Daarnaast biedt deze stap inzicht in wat haalbaar en mogelijk is.

De geformuleerde onderzoeksvragen in dit onderzoek zijn de doelstellingen, het ontwerpen van een gevalideerd maturity model gericht op de gegevenskenmerken: beschikbaarheid en kwaliteit. Het dient toepasbaar te zijn binnen het MKB. Ook zou het MKB-organisaties moeten ondersteunen bij de vraag hoe men kan groeien in volwassenheid op basis van de uitkomsten. Daarvoor worden richtlijnen geleverd (acties) binnen een specifiek focusgebied.

#### Stap 3, *Ontwerp en ontwikkel.*

Deze stap bestaat uit het bepalen van gewenste functionaliteit en het maken van de artefact.

Het ontwerp van het maturity model zal in de kern gebaseerd worden op informatie over maturity modellen uit de onderzochte literatuur. Deze is aangevuld met informatie uit de literatuur over kenmerken van MKB-organisaties. Verschillende typen maturity modellen zullen worden beschouwd in relatie tot MKB-organisatiekenmerken. Op basis daarvan zal een maturity model gekozen en ontworpen worden gericht op toepassing van DA binnen het MKB. Uiteindelijk zal het model, na input van experts en semi gestructureerde interviews, verbeterd worden en getoetst worden in een praktijksituatie.

#### Stap 4, *Demonstratie*.

Dit is de fase van het kwalitatief verbeteren van het artefact. Uit de literatuur blijkt dat het vaak ontbreekt aan validatie van het ontworpen artefact. Door beperkte tijd voor dit onderzoek is een drietal iteraties uitgevoerd.

Door het model te laten beoordelen door experts en feedback te verwerken zal een eerste iteratie worden uitgevoerd. Een tweede iteratie zal plaatsvinden via een semi gestructureerd interview. Feedback zal wederom verwerkt worden en een laatste iteratie zal via een single case study worden uitgevoerd. In hoofdstuk 4 zullen de iteraties verder beschreven worden.

#### Stap 5 De *evaluatie*.

De onderzoeksgegevens die uit de iteratie fases voort komen vormen de basis voor evaluatie van het gehele model. Daarnaast zullen gegevens afkomstig uit het interview en de ontvangen feedback van de experts worden betrokken in de evaluatie.

#### Stap 6, Het *communiceren* van het onderzoek.

Daarbij gaat het om, een verklaring van de toegevoegde waarde, het nut en de effectiviteit aan onderzoekers en andere geïnteresseerden.

Deze communicatie zal in eerste instantie beperkt blijven tot het opleidingsinstituut, experts, case study organisatie en de onderzoeker. Overwogen wordt om na afronding van het volledige onderzoek het rapport te publiceren en delen met de genoemde betrokkenen en geïnteresseerden.

### 3.3. Rigor en Relevance

Hevner et al., (2004) heeft een zevental richtlijnen opgesteld (zie bijlage 9 voor meer details) voor het effectief uitvoeren van DSR onderzoek. Een van de zeven richtlijnen wordt Research Rigor genoemd. Deze richtlijn betreft het op juiste wijze toepassen van bestaande literatuur en onderzoeksmethodes. Een andere richtlijn staat bekend onder de naam Problem Relevance. Deze is gericht op het onderkennen van een relevant probleem. De mate waarin Rigor en Relevance toegepast worden in een onderzoek dient in evenwicht te zijn. Teveel rigor geeft een verlaging van de relevantie en levert daarmee een te beperkte bijdrage aan relevante problematiek. Bij relevantie is juist van belang dat literatuur en onderzoeksmethodes op de juiste wijze worden toegepast. Dat speelt met name bij de ontwerp- (hoofdstuk 2 en 3) en evaluatiefase (hoofdstuk 4).

In de ontwerpfase is bij de literatuurselectie gebruik gemaakt van SLR . Ondanks dat blijft het mogelijk dat een publicatie niet is meegenomen en dat daarmee de (interne) validiteit is beïnvloedt. Het focus area model in dit onderzoek is op basis van bestaande literatuur ontwikkeld. De evaluatie van het onderzoek is beschreven in hoofdstuk 4.

Verschillende factoren kunnen van invloed zijn op de validiteit en de betrouwbaarheid van het onderzoek. In de eerste plaats kan de subjectiviteit van de respondenten de (interne) validiteit beïnvloeden. Daarnaast zal de betrouwbaarheid toenemen naarmate er meer iteraties plaatsvinden. Wanneer namelijk het onderzoek bij herhaling wordt uitgevoerd met telkens een vergelijkbare uitkomst versterkt dat die uitkomst. De rol van de onderzoeker speelt eveneens een belangrijke rol bij de betrouwbaarheid. De interne validiteit en betrouwbaarheid zijn in dit onderzoek geborgd door de interviewvragen uit de tweede iteratie in bijlage 10 toe te voegen. De externe validiteit van dit onderzoek kan worden versterkt doordat het model toepasbaar is binnen andere MKB-organisaties in Nederland. In het buitenland kan dat problematisch zijn vanwege het hanteren van verschillende definities van het begrip MKB-organisatie. Dat leidt tot onduidelijkheid over de vraag wanneer een organisatie een MKB-organisatie is. Daarnaast kunnen verschillen in toepasselijke wet- en



regelgeving leiden tot andere mate van toegankelijkheid van data, waardoor de uitkomsten van onderzoek zich moeilijk laten vergelijken.

In dit onderzoek is gebruik gemaakt van interviews. Er is toestemming verleend voor het maken en na bewerking weer verwijderen van geluidsopnames van deze interviews. Om met respect voor privacy en ethisch verantwoord de resultaten te verwerken zijn organisaties en respondenten in de weergave van de resultaten geanonimiseerd en is gekozen voor respondenten die een onafhankelijke relatie hebben met dit onderzoek.

## 4. Ontwerp, demonstratie, evaluatie en resultaten

In dit hoofdstuk worden voor elke iteratie afzonderlijk het ontwerp, de demonstratie, de evaluatie en de resultaten beschreven. De volledige demonstratie (Peffer's model stap 4) bestond uit een drietal iteraties. Het FAM-model is tijdens de eerste iteratie besproken met acht experts. Deze acht experts zijn elk in het verleden via onderzoek betrokken geweest bij maturity modellen. Enkele van hen waren bij die onderzoeken bovendien (mede) gericht op MKB-organisaties. Input voor de tweede iteratie werd geleverd door drie personen, die werkzaam zijn binnen drie verschillende MKB-organisaties en die bij hun werkzaamheden in die organisatie (onder meer) data-analyse als taak hebben. Via semigestructureerd interview hebben zij hun bijdragen aan dit onderzoek geleverd. De laatste iteratie is via een single case study uitgevoerd. Een persoon die werkzaam is als data analist bij een internationaal opererende MKB-organisatie is bereid gevonden het model binnen zijn organisatie toe te passen.

### 4.1. Iteratie 1

Elke iteratie fase in dit onderzoek heeft bestaan uit een vast aantal onderdelen: ontwerp, demonstratie en evaluatie. De onderdelen van iteratie 1 worden in dit hoofdstuk beschreven.

#### 4.1.1. Ontwerp

##### Keuze maturity model

Bij het kiezen van het type maturity model zijn afwegingen gemaakt op basis van MKB-kenmerken enerzijds en type maturity modellen anderzijds. De in dit onderzoek gebruikte MKB-kenmerken zijn overgenomen van de beschrijving door Kertiwi & MacGregor (2007). Voor type maturity modellen is aangesloten bij het werk van Steenbergen (2011). Zij heeft drie basis types beschreven, staged fixed-level, continuous fixed-level en focus area modellen. Deze modellen zijn in een matrix gezet (bijlage 8) met daarbij vermeld de MKB-kenmerken. Vervolgens is de matrix verder aangevuld met een korte beschrijving per kenmerk.

Gekozen is voor het type Focus Area Maturity model. De belangrijkste kenmerken en bijbehorende beschrijving zijn in Tabel 6 beschreven.

Tabel 6 Belangrijke MKB kenmerken Bron (Kartiwi & MacGregor, 2007)

Kenmerk ID (Kartiwi & MacGregor, 2007)	MKB kenmerk: (Kartiwi & MacGregor, 2007)	Beschrijving:
INT1	SMEs have small and centralized management with a short range perspective.	Het FAM model kent relatief kleinere maturity levels ten opzichte van het Staged fixed level model en het Continuous 5-level model en het kost daarmee kost het relatief minder tijd om een level te voltooien.
INT2	SMEs have poor management skills.	Kleinere maturity stappen implementeren binnen een specifiek focusgebied vereist alleen daarbinnen kennisverdieping. Daarnaast ondersteunen de beschreven aandachtsgebieden hierbij.
INT4	SME owners often withhold information from colleagues	Het FAM maturity model zou door één persoon kunnen worden uitgevoerd.

INT 5	The decision-making process in SMEs is intuitive rather than based on detailed planning and exhaustive study	Een FAM model is eenvoudiger toe te passen en wellicht is het mogelijk om dit door één persoon uit te laten voeren. Daarmee ontstaat meer inzicht bij deze persoon en de mogelijkheid om vervolgens intuïtief beslissingen te nemen.
INT9	SMEs face difficulties obtaining finance and other resources, and as a result have fewer resources	Het FAM model vraagt beperkte kennis en specialisatie. Minder investering is nodig om een volgende maturity fase te voltooien.
INT 11	SMEs have a lack of technical knowledge and specialist staff and provide little IT training for staff	Van de drie modellen vraagt het FAM model de minste kennis bij uitvoering.

### Keuze focusgebieden en dimensies

De onderzoeken van Libes (2015) en Merino et al., (2015) vormen de basis bij de focusgebieden. Samen met het onderzoek van Hardwicke et al., (2018) vormen ze tevens de basis bij de dimensies. Libes (2015) richt zich in zijn onderzoek op de beschikbaarheid van gegevens voor onderzoekers. Het onderzoek van Merino et al., (2016) is gericht op het beoordelen van gegevenskwaliteit in relatie tot Big Data. Aangezien dit onderzoek zich richt op data-analyse binnen MKB-organisaties is elke dimensie (kenmerk) hierboven beoordeeld op toepasbaarheid en waar nodig aangepast. Details over de keuzes en aanpassing per dimensie zijn te vinden in bijlage 11.

Na het doornemen van de literatuur (Hardwicke et al., 2018; Libes, 2015; Merino, et al., 2015) zijn 9 dimensies voor gegevensbeschikbaarheid en 16 dimensies voor gegevenskwaliteit beschreven, zie Tabel 7.

*Tabel 7 Dimensies voor gegevensbeschikbaarheid en gegevenskwaliteit*

<b>Gegevensbeschikbaarheid:</b>	<b>Gegevenskwaliteit:</b>
1. Beperking in interne gegevensgebruik	1. Nauwkeurigheid
2. Beperking in externe gegevensgebruik	2. Volledigheid
3. Niet bestaande of ontbrekende gegevens	3. Consistentie
4. Timing en synchronisatie van gegevens	4. Geloofwaardigheid
5. Gegevensfrequentie	5. Actueel
6. Gegevensformat, standaarden en specificaties	6. Toegankelijkheid
7. Onzekerheid, betrouwbaarheid en nauwkeurigheid	7. Nakoming
8. Toegang tot gegevens, opslag en verwerking	8. Vertrouwelijkheid
9. Voorbereidingstijd van gegevens uit andere bronnen	9. Efficiëntie
	10. Precisie
	11. Traceerbaarheid
	12. Begrijpelijkheid
	13. Beschikbaarheid
	14. Draagbaarheid
	15. Herstelbaarheid
	16. Reproduceerbaar

### Keuze capabilities en maturity levels

In deze fase van het onderzoek ontbrak het aan geschikte literatuur gericht op de vaststelling van capabilities en maturity levels . Level 0 vertegenwoordigt het minimale niveau voor MKB-organisaties (Mittal et al., 2018). Vervolgens is tussen de dimensies een rangorde opgesteld in de mate van belangrijkheid . Dit is zo veel mogelijk gedaan op basis van literatuurinformatie en waar dat niet kon aangevuld vanuit eigen ervaring. Zo ligt bijvoorbeeld voor de hand dat het gebruik van interne informatie eerder zal worden toegepast dan het gebruik van externe informatie.

## Het gehele model

De eerste matrix versie van het maturity model voor gegevensbeschikbaarheid, in tabel 8 weergegeven, is:

Tabel 8 Gegevensbeschikbaarheid matrix

1	Gegevensbeschikbaarheid	0	1	2	3	4	5	6	7	8	9	10	11	12
1.1	Beperking in interne gegevensgebruik	A	B						C			D		
1.2	Beperking in externe gegevensgebruik			A	B				C				D	
1.3	Niet bestaande of ontbrekende gegevens					A	B			C			D	
1.4	Timing en synchronisatie van gegevens			A			B				C			D
1.5	Gegevensfrequentie			A		B	C				D	E		
1.6	Gegevensformat, standaarden en specificaties				A		B	C		D			E	
1.7	Onzekerheid, betrouwbaarheid en nauwkeurigheid				A		B				C			D
1.8	Toegang tot gegevens, opslag en verwerking					A	B		C			D		
1.9	Vorbereidingstijd van gegevens uit andere bronnen							A		B			C	
		DA niet of nauwelijks mogelijk			DA beperkt mogelijk			DA niet volledig mogelijk			DA volledig mogelijk			

Bij de matrix voor gegevenskwaliteit is aan de experts verzocht om aanvullend advies. Vanwege het nog ontbreken van literatuur gericht op het vaststellen van capabilities en maturity levels is gevraagd naar geschikte literatuur. Dit ontbreekt daarom in onderstaande matrix. De enige indeling, overgenomen uit Merino et al., (2015) is naar 'inherente afhankelijke gegevens kwaliteit', 'inherente & systeem afhankelijke gegevens kwaliteit' en 'systeem afhankelijke gegevens kwaliteit'. Deze indeling is overgenomen omdat deze mogelijk ook van belang is in dit onderzoek. Vervolg iteraties zullen dit uitwijzen.

De eerste matrix versie van het maturity model voor gegevenskwaliteit is weergegeven in Tabel 9.

Het ontwikkelde maturity model bestaat uit een aantal dimensies met per dimensie een korte toelichting, een aantal capabilities met daarbij een status beschrijving en een focus beschrijving. De korte toelichting legt uit wat onder deze dimensie wordt verstaan. Het aantal capabilities verschilt per dimensie en is op basis van de geformuleerde stellingen bepaald. Alle stellingen samen dienen het gehele gebied te omvatten dat ligt tussen een totaal ontbreken van volwassenheid en volledige volwassenheid. Daarnaast is bij elke dimensie, per, capability een focus beschrijving toegevoegd die als leidraad kan fungeren wanneer een hogere maturity gewenst is.

2.1 <Dimensie>	
<korte uitleg>	
A: <capability>	<status beschrijving>
Focus:	<Focus / Aandachtsgebied beschrijving>

Tabel 9 Gegevenskwaliteit matrix

<b>2</b>	<b>Gegevenskwaliteit</b>												
	<i>inherente afhankelijke gegevens kwaliteit</i>												
2.1	Nauwkeurigheid - Accuracy		A		B		C		D		E		
2.2	Volledigheid - Completeness		A		B		C		D		E		F
2.3	Consistentie - Consistency		A		B		C		D				
2.4	Geloofwaardigheid - Credibility (Believability)		A		B		C		D		E		
2.5	Actueel - Currentness		A		B		C		D				
	<i>inherente &amp; systeem afhankelijke gegevens kwaliteit</i>												
2.6	Toegankelijkheid - Accesibility		A		B		C		D				
2.7	Nakomen - Compliance		A		B		C		D				
2.8	Vertrouwelijkheid - Confidentiality – data analist		A		B		C		D				
2.9	Vertrouwelijkheid - Confidentiality – informatie behoefte		A		B		C						
2.10	Efficiëntie - Efficiency		A		B		C		D				
2.11	Precisie - Precision		A		B		C		D				
2.12	Traceerbaarheid- Traceability		A		B		C		D		E		
2.13	Begrijpelijkheid - Understandability		A		B		C		D				
	<i>systeem afhankelijke gegevens kwaliteit</i>												
2.14	Beschikbaarheid - Availability		A		B		C		D				
2.15	Transporteerbaar - Portability		A		B		C		D				
2.16	Herstelbaarheid – Recoverability (Raw data)		A		B		C						
2.17	Reproduceerbaar - Reproducible (Analyse proces)		A		B		C		D				

#### 4.1.2. Demonstratie

De eerste iteratie is per email voorgelegd aan de acht eerder genoemde experts. Het maturity model is daarbij aangeboden, voorzien van een inleiding en een verzoek tot het geven van algemene feedback op het model. Het doel daarvan was een aansporing aan de experts om zich niet te beperken tot het beantwoorden van specifieke vragen, maar om opmerkingen en suggesties over het gehele model in de volle breedte te delen met de onderzoeker. Elk van de experts heeft in het verleden onderzoek gedaan naar maturity modellen en bij een enkele van hen was het onderzoek mede gericht op het MKB. Zie in bijlage 12 een tabel met meer informatie over de benaderde experts (geanonimiseerd beschreven als respondenten 1 t/m 8).

Bij de eerste iteratie is er voor gekozen om het geheel op te splitsen in de twee focusgebieden gegevensbeschikbaarheid en gegevenskwaliteit, zodat het te beoordelen geheel per expert relatief beperkt kon blijven. Dat heeft geleid tot twee documenten met een grotendeels vergelijkbare inleiding en een gedeelte met specifieke informatie gericht op het focus area gebied. Aan vier onderzoekers is verzocht de blik te richten op het focusgebied gegevensbeschikbaarheid. Aan de overige vier is aandacht gevraagd voor het focusgebied gegevenskwaliteit. In bijlage 13 staat het FAM-model (eerste deel is gegevensbeschikbaarheid, tweede deel is gegevenskwaliteit) zoals dit is gebruikt bij de eerste iteratie.

### 4.1.3. Evaluatie

Naast de algemene informatie over de benaderde experts staan in bijlage 12 (geanonimiseerd beschreven als respondenten 1 t/m 8) ook hun reacties opgenomen. Van vijf van de acht benaderde experts zijn reacties ontvangen (respons 62,5%). De vijf reacties zijn, in overeenstemming met het doel en de verwachting, algemeen van aard. Opvallend is dat drie van de vijf respondenten opmerkingen maakten over onduidelijkheid van de relatie met het MKB. Aanpassingen die zijn verwerkt op basis van deze eerste iteratie zijn beschreven in hoofdstuk 4.2.1.

## 4.2. Iteratie 2

In elke iteratie-fase is een vast stramien gevolgd, dat bestaat uit een vast aantal onderdelen: ontwerp, demonstratie en evaluatie. In de ontwerp-fase zijn eerst aanpassingen verwerkt op basis van de voorgaande iteratie. Vervolgens is de tweede iteratie uitgevoerd. Tot slot volgden de demonstratie en aansluitend de evaluatie. De onderdelen van iteratie 2 worden in dit hoofdstuk beschreven.

### 4.2.1. Ontwerp

Tijdens de eerste iteratie is feedback ontvangen (zie bijlage 12) van algemene aard is over het gehele model. Dat leverde algemene verbeterpunten op die zich niet lenen voor een concrete beschrijving. Het maturity model is aangepast op basis van de feedback. In dit hoofdstuk zijn de aanpassingen beschreven en het verbeterde model is als basis bij de tweede iteratie gebruikt.

#### **Focusgebieden en dimensies**

De focusgebieden zijn ongewijzigd. De dimensie beschrijvingen zijn op basis de opmerkingen en suggesties uit iteratie 1 herzien. Deze beschrijvingen zijn aangepast en vereenvoudigd.

#### **Capabilities en maturity levels**

De status beschrijvingen per capability zijn heroverwogen naar aanleiding van respondent 7. Deze respondent heeft aanbevolen om de vragen te vereenvoudigen om daarmee de kans op toepasbaarheid te vergroten. Per dimensie zijn in eerste instantie Ja/Nee vragen geformuleerd, waarbij is aangesloten bij het onderzoek van Steenberg (2010). De formulering van Ja/Nee vragen bleek onmogelijk bij een aantal dimensies. Vervolgens is er gekozen voor het formuleren van stellingen. Kenmerkend van stellingen is dat dergelijke vragen makkelijk zijn te beantwoorden omdat er gevraagd wordt welke stelling het best de situatie beschrijft. De kans op onbeantwoord gebleven vragen verkleint daardoor. Dat levert meer informatie op voor elk onderzocht dimensie en vergemakkelijkt de toepassing van het model. Deze vereenvoudiging past goed bij de MKB-kenmerken Int 2, Int 4 en Int 11 (Kartiwi & MacGregor, 2007) zoals beschreven in bijlage 7.

#### **Het gehele model**

Het aangepaste maturity model bestaat na deze fase uit een aantal delen: het inventarisatie deel, de matrix en de aandachtsgebieden. Het inventarisatie deel voor het bepalen van de volwassenheid, de AS-IS situatie per dimensie, bestaat na deze aanpassing uit stellingen. De matrix is ongewijzigd gebleven. Om meer overzicht te verkrijgen zijn de focusbeschrijvingen en de aandachtsgebieden bij elkaar gezet en als laatste onderdeel aan het geheel toegevoegd.

Twee respondenten uit iteratie 1 hadden vraagtekens bij de dimensies zoals deze waren beschreven. Daarnaast werd een opmerking gemaakt over de vragen. Wat eenvoudiger geformuleerde vragen

zou de toepasbaarheid van het model ten goede komen. Naar aanleiding daarvan zijn de beschrijving bij de dimensies nagezien en deels aangepast.

De volgende specifieke vraag werd gesteld aan de experts op het terrein van datakwaliteit: “Binnen het focusgebied datakwaliteit levert de literatuur geen oplossing en zie ik geen onderlinge relaties tussen de beschreven focus area gebieden waardoor het indelen naar onderling verschillende maturity levels voor zover nu zichtbaar voor mij, geen toegevoegde waarde kent. Bent u tegen een vergelijkbare situatie aangelopen, wat is uw advies in deze situatie?” Hierop werd door een expert verwezen naar een onderzoek (Steenbergen et al., 2010) voor een oplossing. Naar aanleiding daarvan zijn de capabilities, overigens pas na de tweede iteratie, aangepast door ze dichter bij elkaar te plaatsen en te groeperen met de kwalificatie “Beperkte gegevens kwaliteit” en “Betere gegevens kwaliteit”.

#### 4.2.2. Demonstratie

De tweede iteratie bestond uit drie semigestructureerde interviews met personen werkzaam in verschillende MKB-organisaties. Deze interviews zijn met toestemming opgenomen en na de geanonimiseerde transcriptie daarvan zijn de opnames verwijderd. De transcripties zijn onder elkaar geplaatst in bijlage 14. Ter voorbereiding op het interview is een aantal telkens terugkerende vragen geformuleerd. Zie voor die vragen Tabel 10 met een indeling naar bruikbaarheid, begrijpelijkheid en gebruikersgemak (Comuzzi & Patel, 2016). Dit geeft structuur en overzicht bij de evaluatie. Vervolgens zijn tijdens het interview de vragen gesteld voor zover daarop nog niet eerder een antwoord was gegeven..

Tabel 10 vooraf opgestelde interview vragen

	Begrijpelijkheid	Gebruikersgemak	Bruikbaarheid
Zijn de dimensies voldoende herkenbaar, waarom wel, waarom niet?	v		
Zijn de dimensies toepasbaar binnen jullie organisatie?			v
Elke dimensie bevat een korte omschrijving. Is die voldoende duidelijk? Zijn er onduidelijke omschrijvingen, zo ja welke en waarom?	v	v	
Elke dimensie bevat vragen, zijn deze voldoende duidelijk? Zijn er momenten waarop je eigenlijk zegt: hier kan ik nauwelijks kiezen?	v	v	v
Zijn er dimensies waarbij je te veel overlap ziet? Welke en hoe zou je dit oplossen, samenvoegen of achterwege laten?	v		
Als we kijken naar de matrix, dan hebben we de score letters. Zijn deze voldoende in te vullen en voldoende duidelijk net als de gegroepeerde mate van DA kwaliteit onderaan de maturity levels.	v	v	
Zijn de aandachtspunten voldoende duidelijk, zijn er onduidelijke omschrijvingen en zo ja welke?	v	v	
Zou deze tool, na aanpassing, naar jouw gevoel geschikt zijn om binnen het MKB toegepast te worden?			v

#### 4.2.3. Evaluatie

De tweede iteratie bestond uit drie semi gestructureerde interviews. Bij een drietal verschillende MKB-organisaties is een interview afgenomen bij een persoon die binnen die organisatie (onder

meer) als taak heeft het uitvoeren van data-analyses, zie Tabel 11. In bijlage 15 is Tabel 11 aangevuld met een kort samenvatting. De drie medewerkers hebben verschillende ervaring en achtergronden.

*Tabel 11 2e iteratie, interview medewerker MKB*

Algemene info
<b>Organisatie 1</b> Online interview via teams Productiebedrijf met 43 medewerkers. Functie geïnterviewde: supply chain manager Opleidingsniveau: MBO Werkervaring: 25 jaar
<b>Organisatie 2</b> Online interview via teams Productiebedrijf met 150 medewerkers. Functie geïnterviewde: data en applicatie beheerder Opleidingsniveau: Universiteit Werkervaring: Half jaar werkzaam bij deze organisatie. (Daarvoor 13,5 jaar als business consultant.)
<b>Organisatie 3</b> Interview op locatie bedrijf Productie bedrijf met 14 medewerkers. Functie geïnterviewde: directeur. Opleidingsniveau: HBO Werkervaring: 22 jaar

Reacties uit de tweede iteratie zijn vervolgens gecodeerd in een Exceldocument. Door toepassing van diverse filters konden de verschillende antwoorden per vraag uit de drie interviews eenvoudig met elkaar worden vergeleken. De antwoorden samen vormen de basis voor de evaluatie. Omdat de tweede iteratie veel gedetailleerde wijzigingen bevat zijn deze wijzigingen in detail beschreven in bijlage 16.

Na feedback van de 2<sup>e</sup> iteratie is het model bijgewerkt. Het daaruit resulterende FAM-model (bijlage 17) is toegepast bij de derde iteratie in de vorm van een single case study.

In de fase van de ontwerp van de eerste iteratie werd de verwachting beschreven dat de drempel voor het gebruik van interne informatie lager zal liggen dan het gebruik van externe informatie. Dat lag voor de hand (zie bladzijde 20). Deze verwachting wordt bevestigd in een van de interviews in de tweede iteratie. De respondent uit organisatie 1 maakte de opmerking “Wij maken geen gebruik van externe data...” terwijl er wel analyses worden gemaakt met gebruikmaking van interne productie data.

### 4.3. Iteratie 3

Elke iteratie fase volgt, zoals beschreven, een vast stramien van een aantal vaste onderdelen: ontwerp, demonstratie en evaluatie. In dit hoofdstuk worden de fasen van iteratie 3 beschreven. Deze iteratie bestond uit een single case study bij een productie- en handelsorganisatie. Het betreft een internationale opererende organisatie met productie in Oost-Europa en verkoop in voornamelijk in West-Europa. Met drie zelfstandige merken wordt in deze onderneming de recreatie markt bediend. De organisatie is marktleider in Nederland en in het Verenigd Koninkrijk. Een



stafffunctionaris die als DA-professional, werkzaam is op het hoofdkantoor van het moederbedrijf, is bereid gevonden het eindmodel toe te passen en te beoordelen op begrijpelijkheid, gebruikersgemak en bruikbaarheid. Ter inleiding is het gehele model kort besproken en toegelicht. Vervolgens is de functionaris begonnen met het toepassen van het model.

### 4.3.1. Ontwerp

De ontwerpbeschrijving voor iteratie 3 is verder verfijnd op basis van de feedback uit iteratie 2.

#### **Inleiding**

Alle drie geïnterviewden bij iteratie 2 waren van mening dat het inleidende deel dat voorafgaat aan het model te weinig informatie bevatte. De inleiding is daarom in zijn geheel herzien en aanzienlijk uitgebreid. De aanvankelijk beknopte inleiding van minder dan een pagina bestaat inmiddels uit drie volle pagina's tekst. Vervolgens zijn enkele afbeeldingen ter verduidelijking toegevoegd en gereviewd door een referent. Het resultaat is te vinden in bijlage 17, de eerste drie pagina's.

#### **Focusgebieden en dimensies**

Een kleine overlapping van dimensies tussen gegevensbeschikbaarheid en gegevenskwaliteit werd niet opgemerkt of als hinderlijk ervaren. De dimensies zijn om die reden niet aangepast. De beschrijving van enkele dimensies leidde tot vragen. Alle beschrijvingen zijn naar aanleiding daarvan opnieuw kritisch bekeken en beoordeeld. De tijdens het interview besproken dimensies zijn herschreven. Respondent 1 gaf bijvoorbeeld een opmerking op dimensie 1.2, beperking in toegang tot externe databronnen. De formulering daarvan is aangepast. Respondent 2 suggereerde dat dit punt mogelijk gesplitst zou kunnen worden naar toegang en gebruik. Deze opmerking is heeft tot heroverweging geleid. Geconcludeerd is dat het vooral gaat om het wel of niet gebruiken van externe data.

#### **Capabilities en maturity levels**

Het onderdeel maturity levels bij gegevenskwaliteit werd als onduidelijk benoemd. Dit heeft geleid tot een aanpassing mede gebaseerd op de ontvangen feedback met verwijzing naar Steenbergen et al. (2012), bij de eerste iteratie. De stellingen werden niet altijd even duidelijk gevonden door de drie respondenten. Sommige stellingen werden als dubbel ervaren: bij een andere stelling gaf respondent 2 aan dat geen keuze kon worden gemaakt aangezien twee stellingen beide evenzeer van toepassing waren op de situatie. De dubbel ervaren stellingen zijn opnieuw kritisch bekeken en beoordeeld. Het aantal stellingen is bij enkele van de dimensies aangepast. De vragen met nummers 1.2, 1.5, 1.6, 2.1, 2.4 zijn bijvoorbeeld met één stelling verminderd, terwijl nummer 2.6 is met één stelling is uitgebreid.

Voor elke stelling is een aandachtsgebied (focus) beschreven. Respondent 2 merkt op dat de focus beschreven bij 1.1 niet aanzette tot actie, terwijl de overige dat wel deden. Dit punt is om die reden herschreven.

#### **Het gehele model**

In het vragengedeelte is de Ja/Nee kolom verwijderd. De Ja/Nee vragen zijn na de eerste iteratie aangepast naar stellingen en daarmee was deze keuzekolom daar niet meer relevant. De beschreven dimensies werden duidelijk bevonden.

Respondent 2 was van mening dat de locatie van de matrix, tussen de stellingen en de aandachtsgebieden in, onpraktisch is. Hij gaf de suggestie om deze bovenaan te plaatsen, waardoor men er eerder op geattendeerd wordt en vervolgens eerder als een antwoordblad zal gebruiken bij

het doornemen van de stellingen. Deze suggestie is overgenomen. De matrix is verplaatst naar het begin, aansluitend achter de inleiding en voorafgaand aan de stellingen.

Verder is een relatie gelegd tussen de kenmerken van MKB-organisaties uit het onderzoek van Kartiwi & MacGregor (2007) (zie bijlage 7) en de focus area's en de dimensies. Dit is in bijlage 18 toegevoegd met daarbij aansluitend de onderliggende argumentatie. Daarmee is in dit onderzoek een verband gelegd tussen de MKB-eigenschappen en de dimensies. In de toekomst zou een herziening van de dimensie beschrijvingen, de bijbehorende stellingen, de matrix en de aandachtsgebieden op kunnen worden uitgevoerd op basis van de gemaakte relatie tussen MKB-kenmerken en de dimensies. Dit is een omvangrijke wijziging op het gehele model zoals dat er nu na iteratie 2 ligt. Vanwege de beperkte onderzoeks-tijd is er voor gekozen dit, in dit afrondende stadium van het onderzoek, buiten beschouwing te laten. Voor dit moment moet worden volstaan met de constatering dat er aanknopingspunten zijn voor vervolgonderzoek.

Naar aanleiding van de feedback uit iteratie 1, gericht op de specifieke vraag of er literatuur is die de keuzes rondom maturity levels en capabilities beschrijft is aanvullende literatuur (Steenbergen et al., 2010) geraadpleegd. Steenbergen et al. (2010) wijst er op dat de indeling van maturity levels en capabilities kan worden ontwikkeld door onderlinge afhankelijkheden te analyseren.

Aangezien we bij gegevenskwaliteit te maken hebben met kenmerken van data, en niet met processen waartussen onderling mogelijk relaties bestaan, bleek het niet mogelijk om een overzicht met onderlinge afhankelijkheden vast te leggen. Wel zijn de lege kolommen met maturity levels verwijderd en zijn de overgebleven maturity levels gegroepeerd in twee categorieën van volwassenheid, beperkte gegevenskwaliteit en betere gegevenskwaliteit.

#### 4.3.2. Demonstratie

In vergelijking tot de demonstratie bij de eerste twee iteraties heeft de derde iteratie niet geleid tot een verbetering van het model op basis van de feedback. Het is een single case study waarbij beoordeeld werd in hoeverre het model nu toepasbaar is binnen een MKB-organisatie.

Het FAM model met daarin verwerkt alle wijzigingen die verzameld zijn tijdens de iteratie fases is voorgelegd aan een data analist (die het model nog niet eerder had gezien, ook niet in een voorgaande versie). Daarbij is toegelicht dat het ongeveer een halfuur tot drie kwartier aan tijd zou vergen. Een meting heeft uitgewezen dat het een uur duurt om de inleiding en de stellingen door te nemen. De matrix is, zonder tussendoor aan de onderzoeker te vragen om verduidelijking, ingevuld. Wel werden twee opmerkingen gemaakt gedurende het doornemen van de stellingen. De eerste opmerking hield in dat het doornemen van de stellingen, en daarmee de dimensies lezend, wel een eyeopener is. Daarmee werd gewezen naar de vele dimensies en het daardoor ontstane besef van wat zoal van belang is in relatie tot DA. De tweede opmerking betrof de omschrijving bij 2.14 'Transporteerbaar'. Deze tekst werd drie keer doorgelezen voordat deze goed te begrijpen werd. Nadat de matrix was ingevuld is aansluitend beoordeeld of er dimensies zijn waarop een hogere volwassenheid gewenst is en waarop dan de aandacht gevestigd zou moeten worden. De aandachtsgebieden werden vervolgens doorgenomen. Na er een paar gelezen te hebben werd de opmerking geplaatst "Dat is handig, vaak is het een beoordeling en is het de vraag 'en nu dan' of 'wat nu dan' maar dit vind ik erg handig". De data analist heeft het ingevulde document na afloop meegenomen. Hij sloot af met de mededeling dat hij er op een rustiger moment nogmaals naar wilde kijken, uit leergierige interesse.

### 4.3.3. Evaluatie

Het ontwikkelde FAM-model is na elke iteratie-fase geëvalueerd. In het DSR-model-model is stap 5 geformuleerd als 'de *evaluatie*, met observeren' daarbij. In dit onderzoek is het observeren toegepast bij het beoordelen van het model. Zonder te vragen wat de reactie van de respondent is en wat de opmerkingen zijn welke geplaatst worden.

### 4.4. Algehele evaluatie

De beschreven dimensies werden zowel bij de iteraties als ook bij de case study als goed ervaren. Hiernaar gevraagd geven de geïnterviewden geen bijzondere opmerkingen dat er nog iets zou ontbreken of dat bepaalde dimensies een te grote mate van overlap zouden hebben, waardoor ze niet als zelfstandige onderdelen of kenmerken kunnen worden gezien. De inleiding is sterk gewijzigd ten opzichte van de eerste versie van het model. Tijdens de iteratie en ook de case study is het essentiële belang van een heldere inleiding voorafgaand aan het model duidelijk geworden. Veel vragen tijdens het afnemen van de interviews hadden bij nader inzien kunnen worden vermeden door heldere informatie in de inleiding. Goede instructies verminderen onduidelijkheden en dragen bij aan interne validiteit (Saunders et al., 2109, p. 516). Een aantal dimensie-omschrijvingen evenals de geformuleerde stellingen werden als lastig ervaren. Bij sommige stellingen kon geen keuze worden gemaakt omdat meerdere opties als best passend werden ervaren. Tijdens de tweede iteratie werd het gehele model nog niet als erg begrijpelijk ervaren. Dat werd mede veroorzaakt door onduidelijke formuleringen of overlappende stellingen, waardoor ook het gebruikersgemak als minder werd ervaren. De data analist in de case study vond het gehele model, oordelend op begrijpelijkheid en gebruiksgemak niet erg eenvoudig maar "wel te doen". Tevens merkte de data analist op, na het gehele model doorgenomen en ingevuld te hebben, nu een beter algeheel beeld van het model was ontstaan. Het kunnen toepassen en de bruikbaarheid werd zeker gezien en onderkend. Het model maakte duidelijk welke dimensies zoal van belang zijn bij DA en dat was volgens de data analist op zichzelf al waardevol. Daarnaast gaf de analist aan dat de meerwaarde vooral zou zitten bij een frequente gebruik van het model.

Het doel van dit onderzoek en de bereikte resultaten worden verder beschreven in hoofdstuk 5.2 de conclusies.

## 5. Discussie, conclusies en aanbevelingen

### 5.1. Discussie

Het model is opgezet op basis van bestaande wetenschappelijke literatuur. Vandaaruit zijn aanpassingen gedaan om het model geschikt te maken voor DA. Bijvoorbeeld binnen het focusgebied gegevensbeschikbaarheid is de dimensie 'gegevens gebruik beperkingen' vervangen door twee nieuwe dimensies: 'beperking in interne gegevensgebruik' en 'beperking in externe gegevensgebruik'. Meerdere onderzoeken gericht op DA beschrijven het gebruik van data en onderscheiden daarbij interne en externe data (Arunachalam et al., 2018; Bavdaz et al., 2011; Immonen et al., 2015; Papachristodoulou et al., 2017). Hierdoor is het beoordelen van de volwassenheid en het opnemen van maatregelen die kunnen leiden tot verbeteringen meer specifiek gemaakt.

Verder is het focusgebied gegevenskwaliteit aangepast. Tijdens de fase van het literatuuronderzoek kwam een extra kenmerk in beeld, dat van reproduceerbaarheid in relatie tot DA (Hardwicke et al., 2018). Kijkende naar de gegevenskwaliteit dimensies was er nog geen dimensie die de reproduceerbaarheid beschreef. Reproduceerbaarheid zegt iets over de mate waarin gegevens kenmerken hebben die het mogelijk maken om een bepaald niveau van bewerkingen en kwaliteit te behouden. Daarbij kan bijvoorbeeld gedacht worden aan toevoeging van metadata bij datasets of de mogelijkheid tot het bewaren van analyse tool instellingen, zodat deze bij herhaalde analyse weer gebruikt kunnen worden. Daarom is de dimensie reproduceerbaarheid toegevoegd aan het maturity model.

De dimensies van gegevensbeschikbaarheid en gegevenskwaliteit zijn overgenomen uit de literatuur van Libes (2015) en Merino et al., (2015) met uitzondering van de dimensie reproduceerbaarheid. Geen van deze onderzoeken beschrijft een praktijk validatie. Door het valideren van het model, met de dimensies uit beide onderzoeken, levert dit onderzoek een bijdrage aan de beoordeling van de betrouwbaarheid van beide onderzoeken.

Met het voltooien van dit onderzoek groeit het inzicht en daarmee de mogelijkheden voor verbeteringen. Een van de doelstellingen was het geschikt maken van een model voor gebruik binnen een MKB-organisatie. Dit doel is naar de mening van de onderzoeker ten dele behaald. Door het meer overzichtelijk maken van het gehele model, het begrijpelijk formuleren van de dimensieomschrijvingen en de stellingen is het model inmiddels toepasbaar. Dit wordt bevestigd tijdens de derde iteratie, een single case study. De bruikbaarheid werd in die fase erkend. Alleen als een individu bij een MKB-organisatie over voldoende specifieke kennis beschikt zou het model toepasbaar zijn. Daarmee is het bij deze stand van het onderzoek nog nauwelijks mogelijk dat MKB-organisaties het als een self service tool gaan toepassen. Wat tijdens de iteratie ook duidelijk werd is dat het model met name voor consultants met klanten in het MKB wel een goede aanvulling zou kunnen zijn, omdat zij zich sneller kunnen bekwamen in het gebruik van het model, door het bij meerdere klanten toe te passen.

Dit onderzoek is gericht op de specifieke kenmerken van gegevens namelijk gegevensbeschikbaarheid en gegevenskwaliteit. Het is niet gericht op de beoordeling van gegevens voor een specifiek proces binnen een bedrijf. Elk proces kan specifieke eisen stellen aan de gebruikte data. Door het universele karakter, het niet specifiek voor één proces ontwikkelde maturity model, kan het model voor verschillende processen worden toegepast. Volwassenheid van gegevens die nodig is voor een specifiek proces, bijvoorbeeld inkoop, zullen anders beoordeeld kunnen worden dan bijvoorbeeld Human Resource data. Het model is generaliseerbaar en kan daarmee bruikbaar worden voor andere processen.

Zoals ook beschreven in hoofdstuk 4.3.1 was het verwerken van de relatie tussen MKB kenmerken en de dimensies gelet op de beschikbare onderzoekstijd niet meer mogelijk. Als de aanpassing wel was doorgevoerd dan was er zonder toetsing in de praktijk alsnog geen sprake van geweest van een gevalideerd model. Daarmee zou deze doelstelling van het onderzoek niet zijn bereikt. Een dergelijke herschrijving zou een herziening inhouden van de dimensie beschrijvingen, de bijbehorende stellingen, de matrix en de aandachtsgebieden.

Hoewel het valideren via een single case study binnen Nederland heeft plaatsgevonden bij een MKB-productie- en handelonderneming is niet uitgesloten dat het model buiten Nederland toepasbaar is en bij ondernemingen die niet actief zijn op het gebied van productie en handel. Dat zal deels afhangen van de lokale wet- en regelgeving ten aanzien van data en de definitie van MKB-organisaties. Ook is niet uitgesloten dat het model toepasbaar is bij grote organisaties. Er is geen relatie tussen gegevenskenmerken (dimensies) en de hoeveelheid gegevens. De dimensies zoals in dit onderzoek beschreven zijn universeel, onafhankelijk van de vraag of men over heel veel gegevens beschikt (grote ondernemingen) of over een beperkte hoeveelheid gegevens (MKB).

## 5.2. Conclusies

Het doel van dit onderzoek was het ontwikkelen van een gevalideerd maturity model dat DA binnen het MKB ondersteund. Het model van Hevner is gevolgd en draagt daarmee bij aan de validiteit. Daarnaast is er een hoofdvraag geformuleerd: Hoe kan een maturity model, gericht op gegevensbeschikbaarheid en gegevenskwaliteit, ontwikkeld worden ter ondersteuning van Data Analytics en dat bovendien toepasbaar is binnen MKB-organisaties?

Om antwoord te krijgen op deze hoofdvraag zijn twee deelvragen geformuleerd.

- Welke eisen kunnen aan gegevensbeschikbaarheid en gegevenskwaliteit worden gesteld voor een MKB-organisatie bij toepassing van Data Analytics?

De eisen die gesteld kunnen worden aan gegevensbeschikbaarheid zijn geformuleerd in de dimensies 1.1 tot en met 1.9, zie bijlage 16. De eisen die gesteld kunnen worden aan gegevenskwaliteit zijn geformuleerd in de dimensies 2.1 tot en met 2.16, zie bijlage 16.

- Welke typen maturity modellen zijn er en waar dient een maturity model gericht op Data Analytics aan te voldoen om bruikbaar te zijn binnen een MKB-organisatie?

In bijlage 8 zijn een drietal maturity modellen (FAM model, Staged 5-level model en Continuous 5-level model) gekoppeld aan MKB kenmerken. Vooral op de kenmerken waar met name kennis en tijd (complexiteit) een rol spelen blijkt het FAM model op basis van de vergelijking het meest geschikt voor het MKB.

Met het antwoord op de deelvragen en terugkerend naar de hoofdvraag in dit onderzoek kan geconcludeerd worden dat op basis van het DSR-model een FAM model ontwikkeld kan worden, gericht op DA en door consultants goed uitvoerbaar binnen MKB-organisaties. Het door een MKB-organisatie zelf uitvoeren van het model blijkt is nog nauwelijks mogelijk door de complexiteit en het ontbreken van de benodigde kennis. Over de vraag of dit in de toekomst misschien wel mogelijk zou zijn kunnen op basis van dit onderzoek geen conclusies worden getrokken.

Wel kan op basis van dit onderzoek geconcludeerd worden dat op basis van het DSR-model (Peffer et al. 2007) een maturity model ontwikkeld kan worden. Het tijdens dit onderzoek ontwikkelde FAM-model is toepasbaar binnen MKB-organisaties. Op basis van de iteraties kan geconcludeerd worden dat het succesvol toepassen van het ontwikkelde model steeds afhankelijk is van de kennis en

affiniteit van degene die het model toepast. Uit de iteratie komt duidelijk naar voren dat het model ideaal is voor gebruik door consultants is die organisaties zoals het MKB ondersteunen en adviseren. Een respondent: "Van de DA welke we hier gebruiken, hebben we het door consultants laten inrichten qua techniek. Wij gebruiken en herhalen hetgeen zij hebben ingesteld.". Het komt er dus op neer dat als een consultant met gebruikmaking van dit model adviezen uitbrengt, het aannemelijk is dat dit toegepast wordt. Een andere respondent beoordeelde het model niet alleen als informatieverwerker maar ook vanuit 13,5 jaar ervaring als business consultant. "Dit is tooling voor een consultant" aldus deze respondent. Consultants kunnen de kwaliteit van hun advies verhogen door tevens o.b.v. een gevalideerd model advies uit te brengen. Tevens kunnen zij de aandachtsgebieden, in dit model zijn het statische beschrijvingen, meer klantspecifiek maken, uitleg geven en begeleiding bieden.

### 5.3. Aanbevelingen voor de praktijk

Het toepassen van het DSR-model van Peffers et al. (2007) zorgt voor goede ontwikkelfases en een goede volgorde van uitvoering. Op basis van de case study word aanbevolen om de inleiding nog te verrijken, door een zin toe te voegen waarin wordt geadviseerd om eerst het model in zijn geheel door te lezen voor een beter begrip en het vervolgens pas toe te passen. Dit geeft meer overzicht dan direct na de inleiding meteen te beginnen met de stellingen.

Zoals het model er nu ligt is het goed toepasbaar voor consultants die zich bezig houden met de ondersteuning van MKB-organisaties. Het model is voor consultants op verschillende manieren toepasbaar. Daarbij kan gedacht worden aan een quick scan om gericht een verbetering te realiseren door stellingen te laten invullen door het MKB bedrijf en zelf met verbetervoorstellen te komen, door toepassing van de aandachtsgebieden, de As-Is situatie te bepalen en de voortgang te meten door het periodiek herhaald uitvoeren van het model.

Aanbevolen wordt om dit onderzoek empirisch te toetsen bij meerdere MKB-organisaties en consultancybedrijven zowel bij grotere als bij kleinere MKB-organisaties, om vast te stellen welke organisatie-kenmerken nodig zijn om te bepalen bij welk type MKB-organisatie de tool goed toepasbaar is. Tevens is het model dermate gewijzigd tussen de tweede iteratie en de single case study dat het opnieuw valideren daarvan tot betere uitkomsten zou kunnen leiden.

### 5.4. Aanbevelingen voor verder onderzoek

Tijdens de iteratie merkte een respondent op dat als men MKB-organisaties in staat wil stellen om zelfstandig (zonder externe kennis) de tool toe te passen overwogen zou moeten worden om een vereenvoudigde versie te ontwikkelen waarbij een afweging wordt gemaakt over de vraag welke dimensies achterwege gelaten kunnen worden of samengevoegd. Dit zou in een empirisch onderzoek verder onder zocht kunnen worden.

Na de derde iteratie, de case study, is er geen aangepast model gemaakt. Een van de aanbevelingen naar aanleiding van deze laatste iteratie is om bij een nieuw model de argumenten zoals beschreven in bijlage 18 in het model te verwerken en de capabilities te herzien en daarmee een meer MKB gericht model te ontwikkelen.

Bij de eerste iteratie is een deel gericht op gegevensbeschikbaarheid. Hierbij heeft 50% van de respondenten feedback gegeven. Met het oog op een betere validiteit te krijgen zou dit nogmaals herhaald kunnen worden onder een grotere groep respondenten.

Tot slot is nog niet onderzocht of het model toepasbaar is bij niet MKB-organisaties. Ook dit zou met vervolgonderzoek getoetst kunnen worden om daarmee zou de generaliseerbaarheid van dit model kunnen worden bepaald.

## Referenties

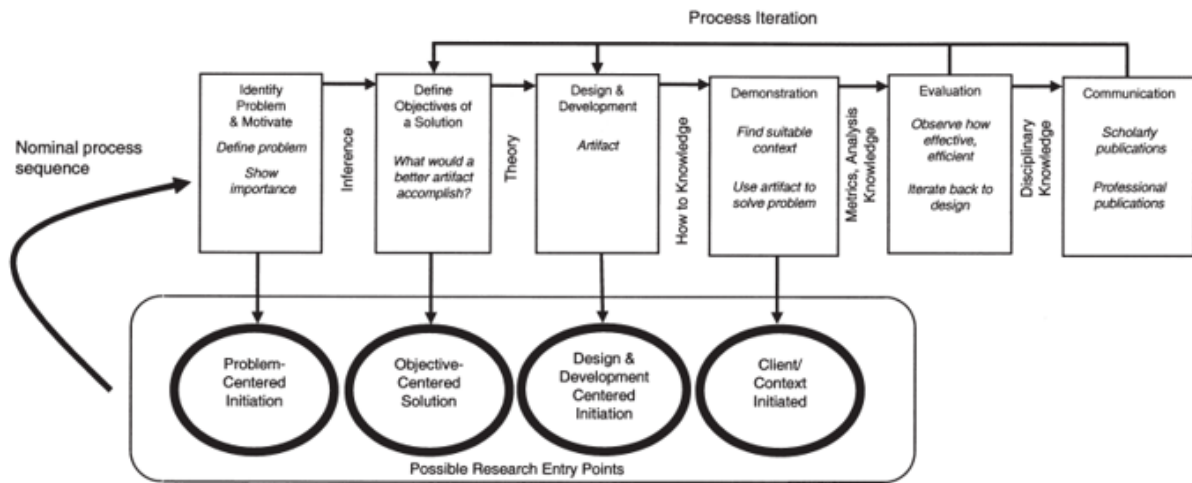
- Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review*, 114, 416–436. <https://doi.org/10.1016/j.tre.2017.04.001>
- Baars, T., Mijnhardt, F., Vlaanderen, K., & Spruit, M. (2016). An analytics approach to adaptive maturity models using organizational characteristics. *Decision Analytics*, 3(1), 1–26. <https://doi.org/10.1186/s40165-016-0022-1>
- Bavdaž, M., Bergström, Y., Biffignandi, S., Bolko, I., Giesen, D., Gravem, D. F., Löfgren, T., Persson, A., Mohorič, P., & Torres van Grinsven, V. (2011). *Business use of NSI statistics based on external sources, Deliverable 3.1. Blue-Ets Project* (Nummer Blue-Ets Project-SSH-CT-2010-244767, Grant Agreement Number 244767).
- Berndtsson, M., M., Forsberg, D., Stein, D., & Svahn, T. (2018). Becoming a data-driven organization. *Conference: European Conference on Information System (ECIS), January*. <https://www.researchgate.net/publication/328233575>
- Bianchini, M., & Michalkova, V. (2019). data analytics in SMEs: trends and policies. *Oecd*, 15, 1–45. <https://doi.org/10.1787/1de6c6a7-en>
- Coleman, S. Y. (2016). Data-mining opportunities for small and medium enterprises with official statistics in the UK. *Journal of Official Statistics*, 32(4), 849–865. <https://doi.org/10.1515/JOS-2016-0044>
- Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., & Reis, M. S. (2016). How Can SMEs Benefit from Big Data? Challenges and a Path Forward. *Quality and Reliability Engineering International*, 32(6), 2151–2164. <https://doi.org/10.1002/qre.2008>
- Comuzzi, M., & Patel, A. (2016). How organisations leverage: Big Data: A maturity model. *Industrial Management and Data Systems*, 116(8), 1468–1492. <https://doi.org/10.1108/IMDS-12-2015-0495>
- Costa, C., Freitas, Â., Stefanik, I., Krafft, T., Pilot, E., Morrison, J., & Santana, P. (2019). Evaluation of data availability on population health indicators at the regional level across the European Union. *Population Health Metrics*, 17(1), 1–15. <https://doi.org/10.1186/s12963-019-0188-6>
- DAMA-NL. (2019). *Code voor Informatiekwaliteit 2019* (p. 124). Stichting DAMA NL, ir Peter W.M. van Nederpelt EMEA EMIA. <http://www.dama-nl.org/wp-content/uploads/2019/09/Code-voor-Informatiekwaliteit-2019-DAMA.pdf>
- Davenport, T. H., Harris, J. G., De Long, D. W., & Jacobson, A. L. (2001). Data to knowledge to results: Building an analytic capability. *California Management Review*, 2, 117–138. <https://doi.org/10.2307/41166078>
- De Bruin, T., Rosemann, M., Freeze, R., & Kulkarni, U. (2005). Understanding the main phases of developing a maturity assessment model. *ACIS 2005 Proceedings - 16th Australasian Conference on Information Systems, January*.
- English, V., & Hoffmann, M. (2018). Business Intelligence as a Source of Competitive Advantage in SMEs: A Systematic Review. *DBS Business Review*, 2, 10–32. <https://doi.org/10.22375/dbr.v2i0.23>
- Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information and Management*, 53(8), 1049–1064. <https://doi.org/10.1016/j.im.2016.07.004>



- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Mohr, A. H., Clayton, E., Yoon, E. J., Tessler, M. H., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, 5(8). <https://doi.org/10.1098/rsos.180448>
- Haug, A., & Arlbjørn, J. S. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, 24(3), 288–303. <https://doi.org/10.1108/17410391111122862>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly: Management Information Systems*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Immonen, A., Pääkkönen, P., & Ovaska, E. (2015). Evaluating the Quality of Social Media Data in Big Data Architecture. *IEEE Access*, 3, 2028–2043. <https://doi.org/10.1109/ACCESS.2015.2490723>
- Juddoo, S. (2015). Overview of data quality challenges in the context of Big Data. *2015 International Conference on Computing, Communication and Security, ICCCS 2015*, December. <https://doi.org/10.1109/CCCS.2015.7374131>
- Kartiwi, M., & MacGregor, R. C. (2007). Electronic Commerce Adoption Barriers in Small to Medium-Sized Enterprises (SMEs) in Developed and Developing Countries. *Journal of Electronic Commerce in Organizations*, 5(3), 35–51. <https://doi.org/10.4018/jeco.2007070103>
- Lazarova-Molnar, S., Mohamed, N., & Al-Jaroodi, J. (2019). Data analytics framework for Industry 4.0: enabling collaboration for added benefits. *IET Collaborative Intelligent Manufacturing*, 1(4), 117–125. <https://doi.org/10.1049/iet-cim.2019.0012>
- Libes, D., Shin, S., & Woo, J. (2015). Considerations and recommendations for data availability for data analytics for manufacturing. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 68–75. <https://doi.org/10.1109/BigData.2015.7363743>
- Liu, Y., Soroka, A., Han, L., Jian, J., & Tang, M. (2020). Cloud-based big data analytics for customer insight-driven design innovation in SMEs. *International Journal of Information Management*, 51 (2020), 102034. <https://doi.org/10.1016/j.ijinfomgt.2019.11.002>
- Melo, P. N., & Machado, C. (2019). Business Intelligence and Analytics in Small and Medium Enterprises. In P. N. M. and C. Machado (Red.), *Business Intelligence and Analytics in Small and Medium Enterprises*. © 2020 by Taylor & Francis Group, LLC. <https://doi.org/10.1201/9780429056482>
- Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A Data Quality in Use model for Big Data. *Future Generation Computer Systems*, 63, 123–130. <https://doi.org/10.1016/j.future.2015.11.024>
- Mijnhardt, F., Baars, T., & Spruit, M. (2016). Organizational characteristics influencing sme information security maturity. *Journal of Computer Information Systems*, 56(2), 106–115. <https://doi.org/10.1080/08874417.2016.1117369>
- Mikalef, P., Pappas, I., Krogstie, J., & Giannakos, M. (2018). Information Systems and e-Business Management. Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and e-Business Management*, 16(3), 547–578.
- Mittal, S., Khan, M. A., Romero, D., & Wuest, T. (2018). A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). *Journal of Manufacturing Systems*, 49(October), 194–214. <https://doi.org/10.1016/j.jmsy.2018.10.005>
- Napoleão, B. M., Felizardo, K. R., De Souza, É. F., & Vijaykumar, N. L. (2017). Practical similarities and differences between Systematic Literature Reviews and Systematic Mappings: A tertiary study.

- Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE, June 2018*, 85–90. <https://doi.org/10.18293/SEKE2017-069>
- Ogbuokiri, Udanor C N, & Agu, M. N. (2015). Implementing big data analytics for small and medium enterprise (SME) regional growth. *IOSR Journal of Computer Engineering Ver. IV*, 17(6), 2278–2661. <https://doi.org/10.9790/0661-17643543>
- Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37(1), 879–910. <https://doi.org/10.17705/1cais.03743>
- Papachristodoulou, E., Koutsaki, M., & Kirkos, E. (2017). Business intelligence and SMEs: Bridging the gap. *Journal of Intelligence Studies in Business*, 7(1), 70–78. <https://doi.org/10.37380/jisib.v7i1.216>
- Parra, X., Tort-Martorell, X., Ruiz-Viñals, C., & Álvarez-Gómez, F. (2019). A maturity model for the information-driven SME. *Journal of Industrial Engineering and Management*, 12(1), 154–175. <https://doi.org/10.3926/jiem.2780>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77.
- Ryu, K. S., Park, J. S., & Park, J. H. (2006). A data quality management maturity model. *ETRI Journal*, 28(2), 191–204. <https://doi.org/10.4218/etrij.06.0105.0026>
- Sandhu, R. S. (1993). *On Five Definitions of Data Integrity* (Nunmer Proc. of the IFIP WG11.3 Workshop on Database Security, p. 11). Department of Information and Software Systems Engineering George Mason University.
- Saunders, M.N.K., Lewis, P., Thornhill, A. (2019). *Research methods for business students, Eighth edition*. Harlow, United Kingdom: Pearson education limited
- Steenbergen, M. (2011). *Maturity and Effectiveness of Enterprise Architecture*. In Utrecht University. <http://igitur-archive.library.uu.nl/dissertations/2011-0609-200519/UUindex.html>
- Steenbergen, M., Bos, R., Brinkkemper, S., Van de Weerd, I., & Bekkers, W. (2010). *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. <https://doi.org/10.1007/978-3-642-13335-0>
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103–110. <https://doi.org/10.1145/253769.253804>
- Vecchio, P. Del, Panniello, U., Minin, A. Di, Petruzzelli, A. M., & Pirri, S. (2017). *Big data for open innovation in SMEs and large corporations : Trends, opportunities, and challenges. December 2016*, 6–22. <https://doi.org/10.1111/caim.12224>
- Venable, J., Pries-heje, J., & Baskerville, R. (2014). *RESEARCH ESSAY FEDS : a Framework for Evaluation in Design Science Research. October 2013*, 1–13. <https://doi.org/10.1057/ejis.2014.36>
- Woodall, P., & Wainman, A. (2015). Measuring Is Service Quality in the Context of the Service Quality -User Satisfaction. *Journal of Information technology Theory and Application*, 7(2), 53–70.

## Bijlage 1, Methodologie van Peffers et al., (2007) DSR-Model



Het DSR-model (Peffers et al., (2007) beschrijft een stapsgewijs onderzoek, een methode gericht op het creëren en evalueren van IT artefacts.

De stappen kort samengevat:

Stap 1, *Probleemidentificatie en motivatie*, omvat het definiëren van het specifieke onderzoeksprobleem, de relevantie en de motivatie of rechtvaardiging van een oplossing. In dit onderzoek word stap 1 uitgewerkt in hoofdstuk 1.

Stap 2, Het afleiden van de *doelstellingen* (kwantitatief of kwalitatief) voor een oplossing uit het onderzoeksprobleem. Daarnaast inzicht in wat haalbaar en mogelijk is. In de volgende stap,

Stap 3, *Ontwerp en ontwikkel*, het bepalen van gewenste functionaliteit en het maken van de artefact, vooraf gegaan door het verzamelen van benodigde theoretische kennis.

Stap 4 is het *testen* van het artefact via een case study, simulatie of experiment.

In stap 5 de *evaluatie*. Meet de resultaten en vergelijk deze met de doelstellingen uit stap 2. Eventueel terugkoppeling naar stap 2. In de laatste stap het *communiceren* van het onderzoek, het aangeven van de toegevoegde waarde, het nut en de effectiviteit aan onderzoekers

## Bijlage 2, Algemene beschrijving Okoli Model

(Systematische Literatuur Review (Okoli et al., 2015)

Het Okoli model kent een viertal onderdelen met per onderdeel een tweetal stappen. De onderdelen zoals Okoli et al., (2015) deze onderscheid zijn: Planning, Selection, Extraction en Execution.

### **Planning (Planning)**

Binnen Planning de stappen literatuurstudie en protocol en training. De literatuurstudie heeft als doel dat er een duidelijk expliciet beeld wordt gecreëerd naar de lezers waarbij het doel en de beoogde doelstellingen identificeerbaar is.

Bij beoordelingen met meer dan één beoordelaar is protocol en training belangrijk. Het is van cruciaal belang dat de beoordelaars volledig duidelijk en in overeenstemming zijn met de details van de procedure welke gevolgd wordt.

### **Selectie (Selection)**

Selectie kent in het model twee onderdelen, Zoeken naar de literatuur en het Practical Screen. Bij literatuur gaat het om een heldere beschrijving van toegepaste literatuur criteria met daarbij beschreven de volledigheid van zoeken. Practical screen beschrijft een expliciete beschrijving van welke literatuur overwogen is ter beoordeling en welke literatuur niet.

### **Samenvatten (Extraction)**

Samenvatten bevat kwaliteitsbeoordeling en data filteren. Alle literatuur welke na de practical screen behouden blijft krijgt bij kwaliteitsbeoordeling en is afhankelijk van de toegepaste onderzoeksmethode per artikel. Daarna zal bij de overgebleven literatuur data gefilterd worden, alleen de gegevens uit de literatuur halen welke relevant is voor het onderzoek.

### **Uitvoering (Excution)**

Analyse en beoordeling zitten in het laatste kenmerk uitvoering. Binnen analyse worden verbanden gelegd en feiten beschreven tussen de verschillende literaturen. Bij de beoordeling wordt op systematische wijze de uitgevoerde methode van literatuuronderzoek beschreven zodanig dat deze onafhankelijk reproduceerbaar is.

## Bijlage 3, Literatuur analyse

Toelichting: 2<sup>e</sup> kolom, "ja" = via snowballing opgenomen. 3<sup>e</sup> kolom, literatuurbron, OU of Scholar. 4<sup>e</sup> kolom, eerste Auteur en publicatie jaar. Rood zie bijlage 7, 6<sup>e</sup> (laatste) kolom is artikel peer reviewed

sno	Bron	Auteur en Jaar van uitgave	Titel	Peer Reviewd
numm	w-balti			
1	OU	Ali, S., et al. (2017).	"Analysis of Interaction between Business Intelligence and SMEs: Learn from Each Other."	ja
2	OU	Boonsritomachai, W., et al. (2016)	"Exploring business intelligence and its depth of maturity in Thai SMEs."	ja
3	OU	Coleman, S. Y. (2016)	"Data-Mining Opportunities for Small and Medium Enterprises with Official Statistics in the UK."	ja
4	OU	Coleman, S., et al. (2016)	"How Can SMEs Benefit from Big Data? Challenges and a Path Forward."	ja
5	OU	Comuzzi, M., et al. (2016)	"How organisations leverage Big Data: a maturity model."	ja
6	OU	Lazarova-Molnar, S., et al. (2019)	"Data analytics framework for Industry 4.0: enabling collaboration for added benefits."	ja
7	ja Scholar	Libes, D., et al. (2015)	Considerations and Recommendations for Data Availability for Data Analytics for	ja
8	OU	Maroufkhani, P., et al. (2020).	"Big data analytics adoption model for small and medium enterprises."	ja
9	OU	Mittal, S., et al. (2018)	"A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small	ja
10	ja Scholar	Baars, T., et al. (2016)	An analytics approach to adaptive maturity models using org characteristics	onbekend
11	OU	Perdana 2020	Getting data analytics on board at The Cage	ja
12	OU	Bach 2019	BPM and BI in SMEs: The role of BPM/BI alignment in organizational performance	ja
13	OU	Del Vecchio, et al., (2016)	Big data for open innovation in SMEs and large corporations: Trends, opportunities, and	ja
14	ja OU	Andriani, M., et al. (2018)	"Aligning business process maturity level with SMEs growth in Indonesian fashion industry."	ja
15	OU	English, V., et al. (2018)	"Business Intelligence as a Source of Competitive Advantage in SMEs: A Systematic Review."	ja
16	OU	Gursu, C. and A. Rancehod (2020).	"Implementing Big Data Analytics in Small Firms: A Situated Human Practice Approach."	ja
17	OU	Liu, Y., et al. (2020)	"Cloud-based big data analytics for customer insight-driven design innovation in SMEs."	ja
18	OU	Papachristodoulou, E., et al. (2017)	"Business intelligence and SMEs: Bridging the gap."	ja
19	ja Scholar	Bavdaz, M., et al. (2011)	Business use of CBS statistics based on external sources	onbekend
20	ja OU	Woodall, P., et al. (2015)	The Data Repurposing Challenge: New Pressure from Data Analytics	ja
21	ja OU	Juddoo, S. (2016)	Data Governance in the Health Industry: Investigating Data Quality Dimensions within a Big Data	ja
22	ja Scholar	Immonen, A., et al. (2015)	Evaluating The Quality of Social Media Data in Big Data Architecture	onbekend
23	ja OU	Hardwicke, T., et al. (2018)	Data availability, reusability, and analytic reproducibility	ja
24	OU	Pirola, F., et al. (2019).	"Digital readiness assessment of Italian SMEs: a case-study research."	ja
25	OU	Gupta, S., et al. (2019).	"Role of cloud ERP and big data on firm performance: a dynamic capability view theory	ja
26	OU	Ingaldi, M. and R. Ulewicz (2019).	"Problems with the Implementation of Industry 4.0 in Enterprises from the SME Sector."	ja
27	OU	Pech, M. and J. Vrchota (2020).	"Classification of Small- and Medium-Sized Enterprises Based on the Level of Industry 4.0	ja
28	OU	Peukert, S., et al. (2020).	4.0 showcases in global production networks."	ja
29	OU	Vitari, C. and E. Raguseo (2019).	"Big data analytics business value and firm performance: linking with environmental context."	ja
30	Scholar	Pereira 2019	Model of Maturity and Audit at SME in Brazil	onbekend
31	OU	Arora 2019	manufacturing using website data	ja
32	OU	Admed 2016	The future of Big Data in facilities management: opportunities and challenges	ja
33	OU	Ahmad, S., et al. (2020).	the Lens of Individual, Technological, Organizational, and Environmental Contexts-A Systematic	ja
34	OU	Batista, L., et al. (2019).	companies."	ja
35	OU	Carballo-Mendivil, B., et al. (2019).	"maturity of processes in smes: validation of scales when implementing project-based learning."	ja
36	OU	Cepel, M., et al. (2018).	"business environment quality index in the sme segment."	ja
37	OU	Dutta, G., et al. (2020).	perspective of Industry 4.0."	ja
38	OU	Ferraris, A., et al. (2019).	"Big data analytics capabilities and knowledge management: impact on firm performance."	ja
39	OU	Fosso Wamba, P. S. (2017).	"Big data analytics and business process innovation."	ja
40	OU	Frederico, G. F., et al. (2019).	"Supply Chain 4.0: concepts, maturity and research agenda."	ja
41	OU	Iqbal, J., et al. (2015).	"Software SMEs' unofficial readiness for CMMI®-based software process improvement."	ja
42	OU	Jun, J.-h., et al. (2020).	"Quality Prediction and Yield Improvement in Process Manufacturing Based on Data Analytics."	ja
43	OU	Kim, E. S., et al. (2019).	SMEs."	ja
44	OU	Lai, Y., et al. (2018).	chain management."	ja
45	OU	Lee, S., et al. (2020).	and Data Mining Techniques."	ja
46	OU	Liu, C., et al. (2019).	and optimal routes planning methods."	ja
47	OU	Maíra Pereira, F. C., et al. (2019).	Enterprises (SMEs) in Brazil."	ja
48	OU	Mantymaki, M., et al. (2020).	of Game Analytics."	ja
49	OU	Maroufkhani, P., et al. (2020).	enterprises."	ja
50	OU	Mittal, S., et al. (2019).	"A smart manufacturing adoption framework for SMEs."	ja
51	OU	Mouzakitis, S., et al. (2016).	analytics."	ja
52	OU	Moyné, J., et al. (2017).	"Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing."	ja
53	OU	O'Connor, C. and S. Kelly (2017).	food sector."	ja
54	OU	Pejić, Bach, M., et al. (2019).	"BPM and BI in SMEs: The role of BPM/BI alignment in organizational performance."	ja
55	OU	Peñá, A., et al. (2018).	"A fuzzy ELECTRE structure methodology to assess big data maturity in healthcare SMEs."	ja
56	OU	Perakis, K., et al. (2020).	scale HPC enabled virtual industrial experimentation environments fostering scalable big data	ja
57	OU	Popovic, A., et al. (2019).	performance."	ja
58	OU	Puklavec, B., et al. (2018).	"Understanding the determinants of business intelligence system adoption stages."	ja
59	OU	Saleem, H., et al. (2020).	product and process innovation matter?"	ja
60	OU	Sanchez-Marquez, R., et al. (2020).	manufacturing sector."	ja
61	OU	Shabbir, M. Q. and S. B. W. Gardezi	knowledge management practices."	ja
62	OU	Singh, A. and S. S. Gill (2020).	capability maturity model integration-based software process improvement."	ja
63	OU	Smith, W. S., et al. (2019).	"Insight from data analytics with an automotive aftermarket SME."	ja
64	OU	Somohano-Rodríguez, F. M., et al.	"Does Industry 4.0 really matter for SME innovation?"	ja
65	OU	Tam, S. and D. E. Gray (2016).	"The practice of employee learning in SME workplaces."	ja
66	OU	Verma, S. and S. S. Bhattacharyya (2016)	qualitative approach for Indian firms."	ja
67	OU	Vickery, A. R., et al. (2019).	Framework."	ja
68	OU	Wang, F., et al. (2020).	"Time Series Data Mining: A Case Study With Big Data Analytics Approach."	ja
69	OU	Widener, D. V., et al. (2018).	intelligence."	ja
70	OU	Zaidan, E. (2017).	countries: The case of Dubai in UAE."	ja
71	Scholar	Triandini 2017	A Maturity Model for E-Commerce adoption by SME in Indonesia	onbekend
72	OU	Steenbergen, M., et al. (2010)	The design of Focus Area Maturity Models	onbekend
73	Scholar	Steenbergen, M. 2011	Maturity and Effectiveness of Enterprise Architecture	onbekend
74	ja Scholar	Mijnhardt, F., (2016)	Organizational Characteristics Influencing SME Information Security Maturity	onbekend
75	ja OU	Merino (2016)	A Data Quality in Use model for Big Data	ja

## Bijlage 4, Resultaat na extraction fase (Data mining resultaat onder groene titels)

na practical v- balli	sno v- nd	Bron	Auteur en Jaar van uitgave	Titel	Peer Reviewd	beschrijft DA	beschrijft MKB	beschrijft Data Kwaliteit	beschrijft Data Kwaliteit	beschrijft Data beschikbaarheid	onderzoekstyp e	Onderzoeks methode	beschrijft Maturity model
1	OU	Boonsitomachai, W., et al. (2016)		"Exploring business intelligence and its depth of maturity in Thai SMEs."	ja	0 niet, 1 matig, 2 ruim	0 niet, 1 matig, 2 ruim	0 niet, 1 matig, 2 ruim	0 niet, 1 matig, 2 ruim	0 niet, 1 matig, 2 ruim	1 (1) kwalitatief	toegepaste Methodolo	0 niet, 1 matig, 2 ruim
2	OU	Coleman, S. Y. (2016)		"Data-Mining Opportunities for Small and Medium Enterprises with Official Statistics in the UK."	ja	2	1	0	0	0	1	monitoring	2
3	OU	Coleman, S., et al. (2016)		"How Can SMEs Benefit from Big Data? Challenges and a Path Forward."	ja	2	1	1	2	2	2	literatuur studie	0
4	OU	Comuzzi, M., et al. (2016)		"How organisations leverage Big Data: a maturity model."	ja	2	1	1	2	2	2	discussie	1
5	OU	Lazarova-Molnar, S., et al. (2019)		"Data analytics framework for Industry 4.0: enabling collaboration for added benefits."	ja	1	0	2	2	2	2	case study	2
6	OU	Libes, D., et al. (2015)		"Considerations and Recommendations for Data Availability for Data Analytics for Manufacturing medium-sized enterprises (SMEs)."	ja	2	1	1	2	2	1	discussie	0
7	OU	Miral, S., et al. (2018)		"A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs)."	ja	1	0	2	2	2	2	discussie	0
8	ja	Scholar	Baars, T., et al. (2016)	"An analytics approach to adaptive maturity models using org characteristics"	onbekend	2	1	0	1	2	2	discussie	2
9	OU	DelVecchio, et al., (2016)		"Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges"	ja	1	0	0	2	2	1	monitoring	2
10	ja	OU	Andriani, M., et al. (2018)	"Aligning business process maturity level with SMEs growth in Indonesian fashion industry."	ja	2	1	0	2	2	2	literatuur studie	0
11	OU	English, V., et al. (2018)		"Business Intelligence as a Source of Competitive Advantage in SMEs: A Systematic Review."	ja	0	1	0	0	0	2	literatuur studie	2
12	OU	Liu, Y., et al. (2020)		"Cloud-based big data analytics for customer insight-driven design innovation in SMEs."	ja	2	1	1	1	2	2	discussie	0
13	OU	Papachristodoulou, E., et al. (2017)		"Cloud-based big data analytics for customer insight-driven design innovation in SMEs."	ja	2	1	0	0	0	2	enquete	0
14	ja	Scholar	Bardaz, M., et al. (2011)	"Business intelligence and SMEs: Bridging the gap."	onbekend	2	1	1	1	1	2	enquete	0
15	ja	OU	Woodall, P., et al. (2015)	"Business use of CBS statistics based on external sources"	ja	1	0	1	1	1	1	interviews	0
16	ja	OU	Juddoo, S. (2016)	"The Data Repurposing Challenge: New Pressure from Data Analytics"	ja	1	0	2	0	0	2	interviews	0
17	ja	Scholar	Immonen, A., et al. (2015)	"Data Governance in the Health Industry: Investigating Data Quality Dimensions within a Big Data Context"	onbekend	1	0	2	1	2	2	discussie	0
18	ja	OU	Hardwicke, T., et al. (2018)	"Evaluating The Quality of Social Media Data in Big Data Architecture"	ja	0	0	2	1	2	2	case study	0
19	ja	Scholar	Steenbergen, M., et al. (2010)	"Data availability, reusability, and analytic reproducibility"	onbekend	0	0	1	2	2	2	literatuur studie	0
20	ja	Scholar	Milhardt, F., (2016)	"The design of Focus Area Maturity Models"	onbekend	0	0	0	0	0	2	literatuur studie	2
				"Organizational Characteristics Influencing SME Information Security Maturity"	onbekend	0	1	0	0	0	2	interviews	2

## Bijlage 5, Libes 2015, databeschikbaarheidsbelemeringen

### Beschikbaarheids issues:

#### A) *Beperking in gegevensgebruik;*

1 Bedrijfseigen gegevens, niet toegankelijk voor derden (en onderzoekers). Daarmee is de bijdrage van onderzoekers tot het leveren van een toepasbare oplossing beperkt.

2 Gegevens beveiliging, ter bescherming van bedrijfsdata, wordt ook binnen de MKB bedrijfsmuren segmentatie van toegang tot data ingericht. Kans op te beperkte data waarop vervolgens Analyse wordt toegepast en dit verkeerde resultaten oplevert.

#### B) *Niet bestaande of ontbrekende gegevens;*

Gegevens kunnen niet verzameld worden vanwege de kosten welke het met zicht meebrengt, tijdelijke storing bij data verzamel opslag of storing bij de sensoren.

#### C) *Timing en synchronisatie van gegevens;*

Bij koppeling van data uit meerdere informatie bronnen is het begrijpen wanneer de informatie is gemaakt belangrijk. De juiste definitie van de vastgelegde tijd is belangrijk om dit te kunnen vergelijken en koppelen met andere data.

#### D) *Gegevensfrequentie;*

De frequentie waarmee gegevens worden verzameld. Een lage frequentie levert minder data op en kan leiden tot minder inzichten. Een hogere frequentie vraagt meer opslag ruimte, kan tot performance problemen leiden en tot hogere kosten.

#### E) *Gegevensformat, standaarden en specificaties;*

Organisaties kunnen uitgebreide verzameling van uiteenlopend gegevens bezitten. Vaak opgeslagen in bestanden geformatteerd volgens verschillende standaarden met een verschil in de mate van nauwkeurigheid.

#### F) *Onzekerheid, betrouwbaarheid en nauwkeurigheid;*

Bedrijfsgegevens kunnen allerlei kenmerken hebben bij aanvang van Data-analyse. Metadata (extra informatie die de data beschrijven) kan ondersteunen daarbij.

#### G) *Toegang tot gegevens, opslag en verwerking;*

Gegevenstoegang en beschikbaarheid zijn van invloed op data-analyse. Zelfs in eenvoudige gevallen worden gegevens structureel gewijzigd wanneer ze uitgedeeld worden aan onderzoekers. Zo is de distributie van digitale gegevens traditioneel een serialisatie (omvormen zodat het transporteerbaar of opslaanbaar is).

#### H) *Vorbereidingstijd voor gegevens en andere bronnen;*

Data verzamelen betekend ook opslag (ruimte) gereed maken voor het bewaren van de data. Welke data waar precies opslaan kost voorbereiding.

## Bijlage 6, Merino 2015, 3 A's Data Kwaliteit-in-gebruik model

<b>Data Quality Characteristic</b>	<b>Contextual Adequacy</b>	<b>Temporal Adequacy</b>	<b>Operational Adequacy</b>
Nauwkeurigheid - Accuracy	X	X	
Volledigheid - Completeness	X		
Consistentie - Consistency	X	X	
Geloofwaardigheid - Credibility	X		
Actueel - Currentness		X	
Toegankelijkheid - Accesibility			X
Nakomen - Compliance	X		
Vertrouwelijk - Confidentiality	X		X
Efficiency - Efficiency			X
Precisie - Precision			X
Traceerbaar – Traceability			X
Begrijpelijkheid - Understandability	X		
Beschikbaarheid – Availability			X
Transporteerbaar - Portability			X
Herstelbaar - Recoverability			X



## Bijlage 7, Kenmerken MKB organisatie, Kartiwi et al., 2007

ID	Features unique to SME	Reported by		
	<i>Features related to management, decision making and planning processes</i>			
INT 1	SMEs have small and centralised management with a short range perspective	Bunker & MacGregor (2000) Welsh & White (1981)	Internal features	
INT 2	SMEs have poor management skills	Blili & Raymond (1993)		
INT 3	SMEs exhibit a strong desire for independence and avoid business ventures which impinge on their independence	Dennis (2000) Reynolds, Savage, & Williams (1994)		
INT 4	SME owners often withhold information from colleagues	Dennis (2000)		
INT 5	The decision-making process in SMEs is intuitive, rather than based on detailed planning and exhaustive study	Reynolds et al. (1994) Bunker & MacGregor (2000)		
INT 6	The SME owner(s) has/have a strong influence in the decision-making process	Reynolds et al. (1994) Bunker & MacGregor (2000)		
INT 7	Intrusion of family values and concerns in decisionmaking processes	Dennis (2000) Bunker & MacGregor (2000) Reynolds et al. (1994)		
INT 8	SMEs have informal and inadequate planning and record keeping processes	Reynolds et al. (1994) Tetteh & Burn (2001) Miller & Besser (2000) Markland (1974) Rotch (1967)		
	<i>Features related to resource acquisition</i>			
INT 9	SMEs face difficulties obtaining finance and other resources, and as a result have fewer resources	Cragg & King (1993) Welsh & White (1981) Gaskill & Gibbs (1994) Reynolds et al. (1994) Blili & Raymond (1993)		
INT 10	SMEs are more reluctant to spend on information technology and therefore have limited use of technology	Walczuch, Van Braven, & Lundgren (2000) Dennis (2000) MacGregor & Bunker (1996) Poon & Swatman (1997) Abell & Limm (1996)		
INT 11	SMEs have a lack of technical knowledge and specialist staff and provide little IT training for staff	Martin & Matlay (2001) Cragg & King (1993) Bunker & MacGregor (2000) Reynolds et al. (1994) Welsh & White (1981) Blili & Raymond (1993)		
	<i>Features related to products/services and markets</i>			
EXT 1	SMEs have a narrow product/service range	Bunker & MacGregor (2000) Reynolds et al. (1994)	External features	
EXT 2	SMEs have a limited share of the market (often confined towards a niche market) and therefore rely heavily on few customers	Hadjimonolis (1999) Lawrence (1997) Quayle (2002) Reynolds et al. (1994)		
EXT 3	SMEs are product oriented, while large businesses are more customer oriented	Reynolds et al. (1994) Bunker & MacGregor (2000) MacGregor, Bunker, & Waugh		

		(1998)	
EXT 4	SMEs are not interested in large shares of the market	Reynolds et al. (1994) MacGregor et al. (1998)	
EXT 5	SMEs are unable to compete with their larger counterparts	Lawrence (1997)	
	<i>Features related to risk taking and dealing with uncertainty</i>		
EXT 6	SMEs have lower control over their external environment than larger businesses, and therefore face more uncertainty	Westhead & Storey (1996) Hill & Stewart (2000)	
EXT 7	SMEs face more risks than large businesses because the failure rates of SMEs are higher	Brigham & Smith (1967) DeLone (1988) Cochran (1981)	
EXT 8	SMEs are more reluctant to take risks	Walczuch et al. (2000) Dennis (2000)	

## Bijlage 8, MKB kenmerken in relatie tot type maturity model

	type Maturity model MKB kenmerk	FAM model (FAM)	Staged 5-level model (S5)	Continuous 5-level model (C5)
INT 1	SMEs have small and centralised management with a short range perspective	Het FAM model kent relatief kleinere maturity levels ten opzichte van S5 en C5 en daarmee relatief minder tijd kost om een level te vervullen.	Het S5 model kent grotere gefaseerde maturity levels welke elk relatief meer tijd kost om te vervullen.	Het C5 model kent grotere gefaseerde maturity levels welke elk relatief meer tijd kost om te vervullen.
INT 2	SMEs have poor management skills	Kleinere maturity stappen implementeren binnen een specifieke focusgebied vereist alleen daarbinnen kennis verdieping.	Grote maturity stappen implementeren vereist diepere kennis van de diverse onderdelen.	Het continue nemen van vervolg maturity stappen vereist diepere kennis en blijvende aandacht.
INT 3	SMEs exhibit a strong desire for independence and avoid business ventures which impinge on their independence	Het FAM maturity model beïnvloed de onafhankelijkheid van de onderneming niet.	Het S5 maturity model beïnvloed de onafhankelijkheid van de onderneming niet.	Het C5 maturity model beïnvloed de onafhankelijkheid van de onderneming niet.
INT 5	The decision-making process in SMEs is intuitive rather than based on detailed planning and exhaustive study	Van de 3 typen maturity modellen vraagt de FAM de minste kennis voor uitvoering. Als er een model is wat een bijdrage kan leveren aan DA maturity dan zou dat het FAM model zijn.	Van de 3 typen maturity modellen vraagt de S5 kennis voor uitvoering. Het is daarmee minder geschikt voor toepassing.	Van de 3 typen maturity modellen vraagt de C5 kennis voor uitvoering. Het is daarmee minder geschikt voor toepassing.
INT 4, 6 t/ m 8		Het FAM maturity model beïnvloed dit kenmerk niet of nauwelijks.	Het S5 maturity model beïnvloed dit kenmerk niet of nauwelijks.	Het C5 maturity model beïnvloed dit kenmerk niet of nauwelijks.
	<b>Features related to resource acquisition</b>			
INT 9	SMEs face difficulties obtaining finance and other resources, and as a result have fewer resources	Het FAM model vraagt beperkte kennis en specialisatie. Minder investering is nodig om een volgende maturity fase te realiseren.	Het S5 maturity model vraagt kennis en specialisatie. Als snel is daar een investering mee gemoeid om maturity fases te realiseren.	Het C5 maturity model vraagt kennis en specialisatie. Als snel is daar een investering mee gemoeid om maturity fases te realiseren.
INT 10	SMEs are more reluctant to spend on information technology and therefore have limited use of technology	Het FAM maturity model zelf beïnvloed niet de mate van technologie gebruik.	Het S5 maturity model zelf beïnvloed niet de mate van technologie gebruik.	Het C5 maturity model zelf beïnvloed niet de mate van technologie gebruik.
INT 11	SMEs have a lack of technical knowledge and specialist staff and provide little IT training for staff	Van de drie modellen vraagt het FAM model de minste kennis bij uitvoering.	Van de drie modellen vraagt het S5 model de nodige kennis bij uitvoering.	Van de drie modellen vraagt het C5 model de nodige kennis bij uitvoering.
	<b>Features related to products/services and markets</b>			
EXT 1 t/m 8		Het FAM maturity model beïnvloed dit kenmerk niet of nauwelijks.	Het S5 maturity model beïnvloed dit kenmerk niet of nauwelijks.	Het C5 maturity model beïnvloed dit kenmerk niet of nauwelijks.

## Bijlage 9, De zeven richtlijnen van Hevner (2004), p. 12

De zeven richtlijnen van Hevner, aangevuld met de omschrijving zoals beschreven in zijn onderzoek.

### Guideline 1: Design as an Artifact

Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.

### Guideline 2: Problem Relevance

The objective of design-science research is to develop technology-based solutions to important and relevant business problems.

### Guideline 3: Design Evaluation

The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.

### Guideline 4: Research Contributions

Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.

### Guideline 5: Research Rigor

Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.

### Guideline 6: Design as a Search Process

The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.

### Guideline 7: Communication of Research

Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

## Bijlage 10, Interview vragen uit de 2e iteratie

De onderstaande vragen zijn vooraf aan de tweede iteratie ronde opgesteld. De vragen zijn expliciet gesteld zodra deze niet tijdens het interview al reeds beantwoord waren.

### Vragen gericht op de dimensies, omschrijvingen en de stellingen .

1 *Zijn de dimensies voldoende herkenbaar*, waarom wel, waarom niet?

2 *Zijn de dimensies* toepasbaar binnen jullie organisatie?

3 Elke dimensie bevat een korte omschrijvingen. Is dat voldoende duidelijk? Zijn er onduidelijke omschrijvingen, zo ja welke en waarom?

4 Elke dimensie bevat vragen, zijn deze voldoende duidelijk? Zijn er moment waarop je eigenlijk zegt hier kan in nauwelijks kiezen?

5 zijn er dimensies waarbij je te veel overlap ziet? Welke en hoe zou je dit oplossen, samenvoegen of achterwege laten?

### Vragen gericht op het Matrix model

6 Als we kijken naar de matrix, dan hebben we de score letters. Zijn deze voldoende in te vullen en voldoende duidelijk net als de gegroepeerde mate van DA kwaliteit onderaan de maturity levels.

### Vragen gericht op de aandachtspunten

7 Zijn de aandachtspunten voldoende duidelijk, zijn er onduidelijke omschrijvingen zo ja welke?

### Vragen gericht op het algeheel toepasbaarheid binnen het MKB

8 Zou deze tool, naar aanpassing, naar jouw gevoel geschikt zijn om binnen het MKB toegepast te worden?

## Bijlage 11, Artefact Dimensie keuzes

### Gegevensbeschikbaarheid

Dimensie: Gegevens gebruik beperkingen
<p>Beschrijving:</p> <ol style="list-style-type: none"> <li>1 Bedrijfseigen gegevens, niet toegankelijk voor derden (en onderzoekers). Daarmee is de bijdrage van onderzoekers tot het leveren van een toepasbare oplossing beperkt.</li> <li>2 Gegevens beveiliging, ter bescherming van bedrijfsdata, wordt ook binnen de MKB bedrijfsmuren segmentatie van toegang tot data ingericht. Kans op te beperkte data waarop vervolgens Analyse wordt toegepast en dit verkeerde resultaten oplevert.</li> </ol>
<p>Beoordeling: Diverse literatuur (McGrath &amp; Burgess, 2016; Immonen et al., 2015; Woodall &amp; Wainman, 2015)) beschrijven gegevens naar interne en externe gegevens bij Data Analytics. Voor het beoordeling van gegevens gebruik zal dit onderzoek een specifiek onderscheid worden gemaakt naar intern en externe gegevens. Deze dimensie is gesplitst en een aanpassing in de omschrijving.</p>
<p>Dit kenmerk in het maturity model als volgt beschreven: (dimensie en omschrijving)</p> <ol style="list-style-type: none"> <li>1 Beperking in interne gegevensgebruik. In welke mate heeft data analist toegang tot benodigde interne data.</li> <li>2 Beperking in externe gegevensgebruik. In welke mate beschikt de data analist over externe data of kan deze er over beschikken.</li> </ol>

Dimensie: Niet bestaande of ontbrekende data
<p>Beschrijving: Gegevens kunnen niet verzameld worden vanwege de kosten welke het met zicht meebrengt, tijdelijke storing bij data verzamel opslag of storing bij de sensoren.</p>
<p>Beoordeling: Een algemene omschrijving welke in het model meer gericht op DA is beschreven.</p>
<p>Dit kenmerk in het maturity model als volgt beschreven:</p> <p>In hoeverre wordt er voor specifieke DA nog extra data verzameld binnen de organisatie (interne data), naast de normale proces data, om een meer gefundeerdere analyse te kunnen uitvoeren.</p>

Dimensie: Timing en synchronisatie van gegevens
<p>Beschrijving: Bij koppeling van data uit meerdere informatie bronnen is het begrijpen wanneer de informatie is gemaakt belangrijk. De juiste definitie van de vastgelegde tijd is belangrijk om dit te kunnen vergelijken en koppelen met andere data.</p>
<p>Beoordeling: Een algemene omschrijving welke in het model meer gericht op DA is beschreven.</p>
<p>Dit kenmerk in het maturity model als volgt beschreven:</p> <p>Bij data-analyse kan het moment (tijd) waarbij een waarde is vastgelegd belangrijk zijn. Vooral bij het gebruik van meerdere data bronnen is het van belang dat er zekerheid is zodat geregistreerde tijden onderling vergelijkbaar zijn bijvoorbeeld chronologische volgorde van vastgelegde data.</p>

Dimensie: Gegevensfrequentie
<p>Beschrijving: De frequentie waarmee gegevens worden verzameld. Een lage frequentie levert minder data op en kan leiden tot minder inzichten. Een hogere frequentie vraagt meer opslag ruimte, kan tot performance problemen leiden en tot hogere kosten.</p>
<p>Beoordeling: Een algemene omschrijving welke in het model meer gericht op DA is beschreven.</p>
<p>Dit kenmerk in het maturity model als volgt beschreven:</p>

De frequentie waarop interne data wordt verzameld / vastgelegd. (bijvoorbeeld bij sensoren op productiemachines). Meer beschikbare data kan leiden tot hogere nauwkeurigheid en betere inzichten, meer opslag betekent o.a. meer opslagruimte vereist en hogere kosten.

**Dimensie: Gegevensformat, standaarden en specificaties**

**Beschrijving:** Organisaties kunnen uitgebreide verzameling van uiteenlopend gegevens bezitten. Vaak opgeslagen in bestanden geformatteerd volgens verschillende standaarden met een verschil in de mate van nauwkeurigheid.

**Beoordeling:** Een algemene omschrijving welke in het model meer gericht op DA is beschreven.

Dit kenmerk in het maturity model als volgt beschreven:

Organisaties kennen ongelijksoortige gegevens en vaak opgeslagen in bestanden geformatteerd volgens verschillende normen.

Het kunnen normaliseren / standaardiseren van vast te leggen data voor analyse kan van belang zijn.

**Dimensie: Onzekerheid, betrouwbaarheid en nauwkeurigheid**

**Beschrijving:** Bedrijfsgegevens kunnen allerlei kenmerken hebben bij aanvang van Data-analyse. Metadata (extra informatie die de data beschrijven) kan ondersteunen daarbij.

**Beoordeling:** Een algemene omschrijving welke in het model meer gericht op DA is beschreven.

Dit kenmerk in het maturity model als volgt beschreven:

Bedrijfsgegevens kunnen kenmerken hebben welke van invloed zijn op Data Analytics zoals onzekerheid, betrouwbaarheid en nauwkeurigheid. Metadata kan hier bij helpen.

Het feit dat analyses in toenemende mate bepalend zijn voor besluitvorming, is het van essentieel belang dat deze data betrouwbaar is.

**Dimensie: Toegang tot gegevens, opslag en verwerking**

**Beschrijving:** Gegevenstoegang en beschikbaarheid zijn van invloed op data-analyse. Zelfs in eenvoudige gevallen worden gegevens structureel gewijzigd wanneer ze uitgedeeld worden aan onderzoekers. Zo is de distributie van digitale gegevens traditioneel een serialisatie (omvormen zodat het transporteerbaar of opslaanbaar is).

**Beoordeling:** Een algemene omschrijving welke in het model meer gericht op DA is beschreven.

Dit kenmerk in het maturity model als volgt beschreven:

Bij toegang is de vraag of er direct op de productie data gewerkt wordt of dat er voor DA met een kopie gewerkt wordt. Bij dit laatste kan data onbewust aangepast zijn zodat het als kopie opgeslagen kan worden, denk een grote dataset waar alleen maar een klein deel nodig is. Daarnaast kan data in verschillende type databases opgeslagen zijn. Elke heeft zijn voor- nadelen bij verwerking in DA.

**Dimensie: Voorbereidingstijd voor gegevens en andere bronnen**

**Beschrijving:** Data verzamelen betekend ook opslag (ruimte) gereed maken voor het bewaren van de data. Welke data waar precies opslaan kost voorbereiding.

**Beoordeling:** Een algemene omschrijving welke in het model meer gericht op DA is beschreven.

Dit kenmerk in het maturity model als volgt beschreven:

Het verkrijgen van data uit andere (intern / extern) niet standaard bronnen vereist voorbereiding voor verwerking. Voor verwerking en interpretatie dient de ontvangen data geanalyseerd te worden voorafgaand aan verdere verwerking binnen DA en eventuele metadata gegevens doorgenomen te worden.

## Gegevenskwaliteit

Merino et al. (2016) geven zelf geen beschrijving van de dimensie in hun onderzoek. Zij verwijzen naar ISO 25012 normering. Onderstaande beschrijvingen zijn uit ISO 25012 overgenomen.

Dimensie: Nauwkeurigheid
Beschrijving: (ISO25012) De mate waarin de gegevens attributen hebben die de werkelijke waarde van het beoogde attribuut van een concept of gebeurtenis in een specifieke gebruikscontext correct weergeven.
Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
Dit kenmerk in het maturity model als volgt beschreven: De nauwkeurigheid geeft aan hoeveel de meetwaarde of informatie van de werkelijkheid af zit. Nauwkeurigheid is te bepalen door de meting te vergelijken met een meting waarvan de nauwkeurigheid bekend is.

Dimensie: Volledigheid
Beschrijving: (ISO25012) De mate waarin subjectgegevens die aan een entiteit zijn gekoppeld, waarden hebben voor alle verwachte attributen en gerelateerde entiteitsinstanties in een specifieke gebruikscontext.
Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
Dit kenmerk in het maturity model als volgt beschreven: Bepalen welke datavelden je nodig hebt voor de analyse. Daarnaast per dataveld controle of het daadwerkelijk gegevens bevat.

Dimensie: Consistentie
Beschrijving: (ISO25012) De mate waarin gegevens attributen hebben die vrij zijn van tegenstrijdigheden en coherent zijn met andere gegevens in een specifieke gebruikscontext.
Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
Dit kenmerk in het maturity model als volgt beschreven: De gegevens mogen geen hiaten bevatten in het tijdspad van analyse (tijdconsistentie), geen data duplicaten en vrij van tegenstrijdigheden. Hebben meet sensoren bijvoorbeeld continue goed gefunctioneerd en data constant aangeleverd, zijn handmatige formulieren door iedereen op een zelfde wijze ingevuld.

Dimensie: Geloofwaardigheid
Beschrijving: (ISO25012) De mate waarin gegevens attributen hebben die door gebruikers in een specifieke gebruikscontext als waar en geloofwaardig worden beschouwd.
Beoordeling: Een algemene omschrijving welke nagenoeg volledig is overgenomen.
Dit kenmerk in het maturity model als volgt beschreven: De mate waarin gegevens kenmerken hebben die in een specifieke gebruikscontext als waar en geloofwaardig (authentiek) kunnen worden beschouwd en de verzameling op een juiste wijze heeft plaatsgevonden.



Dimensie: Actueel
Beschrijving: (ISO25012) De mate waarin data attributen hebben die van de juiste leeftijd zijn in een specifieke gebruikscontext.
Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
Dit kenmerk in het maturity model als volgt beschreven: Bevat de data kenmerken van de juiste leeftijd binnen een specifieke analyse context.

Dimensie: Toegankelijkheid
Beschrijving: (ISO25012) De mate waarin gegevens toegankelijk zijn in een specifieke gebruikscontext, met name: door mensen die vanwege een handicap ondersteunende technologie of een speciale configuratie nodig hebben
Beoordeling: Een algemene omschrijving welke nagenoeg volledig is overgenomen.
Dit kenmerk in het maturity model als volgt beschreven: De mate waarin gegevens toegankelijk zijn in een specifieke gebruikscontext, met name door mensen die ondersteunende technologie of speciale configuratie nodig hebben vanwege een toegangsbeperking.

Dimensie: Naleven
Beschrijving: (ISO25012) De mate waarin gegevens attributen hebben die voldoen aan standaarden, geldende conventies of regelgeving en soortgelijke regels met betrekking tot datakwaliteit in een specifieke context bruikbaar.
Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
Dit kenmerk in het maturity model als volgt beschreven: De mate waarin gegevens kenmerken hebben die voldoen aan geldende normen, conventies of voorschriften en soortgelijke regels met betrekking tot gegevenskwaliteit in een specifieke gebruikscontext.

Dimensie: Vertrouwelijkheid
Beschrijving: (ISO25012) Het is de mate waarin gegevens attributen hebben die ervoor zorgen dat ze alleen toegankelijk zijn en interpreteerbaar door geautoriseerde gebruikers in een specifieke gebruikscontext.
Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
Dit kenmerk in het maturity model als volgt beschreven: Hoe vertrouwelijk is de data binnen de organisatie. In welke mate is het toegankelijk en is deze interpreteerbaar zonder kennis van de context.

Dimensie: Efficiëntie
Beschrijving: (ISO25012) Het is de mate waarin gegevens attributen hebben die kunnen worden verwerkt en die de verwachte prestatieniveaus door de juiste hoeveelheden en soorten middelen te gebruiken in een specifieke gebruikscontext.
Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
Dit kenmerk in het maturity model als volgt beschreven:

Het kunnen aanpassen van data zodat dit de prestatie van analyse verbetert bijvoorbeeld het verkleinen van de data omvang (deductie, compressie, ontdebellen)
---------------------------------------------------------------------------------------------------------------------------------------------------------------

Dimensie: Precisie
--------------------

Beschrijving: (ISO25012)
--------------------------

De mate waarin gegevens attributen hebben die exact zijn of discriminatie opleveren in een specifieke gebruikscontext.
------------------------------------------------------------------------------------------------------------------------

Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
--------------------------------------------------------------------------------------------

Dit kenmerk in het maturity model als volgt beschreven:
---------------------------------------------------------

In hoeverre bevat de data waardes welke, afhankelijk van de context, op meerdere wijzen interpreteren baar is.
----------------------------------------------------------------------------------------------------------------

Dimensie: Traceerbaarheid
---------------------------

Beschrijving: (ISO25012)
--------------------------

De mate waarin gegevens attributen hebben die een controlespoor bieden voor toegang tot de gegevens en van eventuele wijzigingen in de gegevens in een specifieke gebruikscontext.
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
--------------------------------------------------------------------------------------------

Dit kenmerk in het maturity model als volgt beschreven:
---------------------------------------------------------

Kunnen we controleren op mutaties in de data nadat deze data is gebruikt voor analyse. Zijn we in staat om een analyse nogmaals toe te passen op de data en is de uitkomst gelijk.
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Dimensie: Begrijpelijkheid
----------------------------

Beschrijving: (ISO25012)
--------------------------

De mate waarin gegevens attributen hebben waarmee ze kunnen worden gelezen en geïnterpreteerd door gebruikers, en worden uitgedrukt in de juiste talen, symbolen en eenheden in een specifieke context van gebruik.
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
--------------------------------------------------------------------------------------------

Dit kenmerk in het maturity model als volgt beschreven:
---------------------------------------------------------

In hoeverre hebben gegevens kenmerken die het mogelijk maken om door gebruikers te lezen en te interpreteren, en worden uitgedrukt in geschikte talen, symbolen en eenheden in een specifieke gebruikscontext. Sommige informatie beschikt over metadata, een begeleidend schrijven.
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Dimensie: Beschikbaarheid
---------------------------

Beschrijving: (ISO25012)
--------------------------

De mate waarin gegevens attributen hebben waarmee ze kunnen worden opgehaald door geautoriseerde gebruikers en/of applicaties in een specifieke gebruikscontext.
------------------------------------------------------------------------------------------------------------------------------------------------------------------

Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.
--------------------------------------------------------------------------------------------

Dit kenmerk in het maturity model als volgt beschreven:
---------------------------------------------------------

In hoeverre is de data direct gereed of is preparatie nodig om geanalyseerd te kunnen worden.
-----------------------------------------------------------------------------------------------

Dimensie: Transporteerbaarheid
--------------------------------

Beschrijving: (ISO25012)
--------------------------

De mate waarin gegevens attributen hebben waarmee ze kunnen worden geïnstalleerd, vervangen of verplaatst van het ene systeem naar het andere met behoud van de bestaande kwaliteit in een specifieke gebruikscontext.
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Beoordeling: Een algemene omschrijving welke nagenoeg volledig is overgenomen.
--------------------------------------------------------------------------------

Dit kenmerk in het maturity model als volgt beschreven:  
 De mate waarin gegevens kenmerken hebben waarmee deze kunnen worden geïnstalleerd, vervangen of verplaatst van het ene systeem naar het andere, met behoud van de bestaande kwaliteit in een specifieke gebruikscontext.

Dimensie: Herstelbaarheid

Beschrijving: (ISO25012)

De mate waarin gegevens attributen hebben waarmee ze kunnen worden onderhouden en bewaard een bepaald niveau van bedrijfsvoering en kwaliteit, ook in geval van storing, in een specifieke context van gebruik.

Beoordeling: Een algemene omschrijving welke in het model wat vereenvoudigd is beschreven.

Dit kenmerk in het maturity model als volgt beschreven:

De mate waarin ruwe data kan worden hersteld (terug halen van backup) naar de situatie waarop analyse heeft plaats gevonden en exact dezelfde analyse weer uitgevoerd kan worden.

Dimensie: Reproduceerbaar

Beschrijving: Toegevoegd op basis van informatie uit het onderzoek van Hardwick et al. (2018), zie beoordeling.

Het is de mate waarin de gegevens ook herbruikbaar zijn en bij eenzelfde analyse dezelfde uitkomsten oplevert.

Beoordeling: Hardwick et al. (2018) beschrijft een aantal onderzoeken waarin reproduceerbaarheid van data-analyse is onderzocht. Centraal in die onderzoeken is het kunnen reproduceren van onderzoek analyses met behulp van dezelfde data. Bij data-analyse is het van belang dat als er een analyse is gedaan deze analyse bij herhaling exact dezelfde uitkomsten geeft. Aangezien de overige dimensies niet of nauwelijks dit kenmerk beschrijven dit kenmerk toegevoegd.

Dit kenmerk in het maturity model als volgt beschreven:

De mate waarin gegevens kenmerken hebben die het mogelijk maken om een bepaald niveau van bewerkingen en kwaliteit te behouden, zelfs in het geval van een storing, in een specifieke gebruikscontext (vastleggen van bijvoorbeeld gebruikte query, metadata beschrijving, eventuele gemaakte specifieke instellingen binnen een tool).

## Bijlage 12, 1<sup>e</sup> iteratie, benaderde experts en feedback

Geanonimiseerd overzicht van benaderde onderzoekers en hun reactie.

Een aantal benaderd over gegevensbeschikbaarheid en een aantal over gegevenskwaliteit

Per email benaderd en een document verstuurd. In dat document, voorafgaand aan het model een korte beschrijving van het onderzoek, de onderzoeksvragen, het concept maturity model en de context van dit onderzoek. Aansluitend de gehanteerde definitie van gegevensbeschikbaarheid resp. datakwaliteit.

<b>Gegevenskwaliteit</b>	
<b>Respondent/expert (geanonimiseerd)</b>	<b>Feedback</b>
<b>Respondent 1</b> Universitair docent  Meerdere publicaties laatste 15 jaar o.a. gericht op Business Proces management en Analytics, Informatica.  Meer dan 90 publicaties	Geen reactie
<b>Respondent 2</b> Universitair docent  Meerdere publicaties laatste 18 jaar o.a. gericht op Informatie en computer wetenschap, Organisatie en Informatie.  Meer dan 400 publicaties	Geen reactie
<b>Respondent 3</b> Universitair docent  Meerdere publicaties laatste 13 jaar o.a. gericht op Informatica, Intelligent Software Systems  Meer dan 180 publicaties	-interessant probleem -op dit vlak ontbreekt nog gevalideerde kennis -nog niet voldoende duidelijk relatie met model en MKB -Beargumenteerde aanpassingen o.b.v. Merino 2015.
<b>Respondent 4</b> Onderzoeker aan universiteit  Onderzoeken gericht op Mathematics and Computer	-aantal dimensies vermoedelijk lastig binnen MKB -Beschrijving van vragen meer helder maken, deze zijn te onduidelijk.

Science, Cyber Security en Maturity bij MKB	
<b>Gegevensbeschikbaarheid</b>	
<b>Respondent 5</b> Lector aan universiteit en consultant bij IT organisatie Meerdere publicaties laatste 10 jaar o.a. gericht op IT architectuur	Geen reactie
<b>Respondent 6</b> Lector aan hogeschool  Meerdere publicaties laatste 5 jaar met focus op data groei en toepassing	-interpretatie van enkele dimensie aanpassen, wellicht meer MKB specifiek maken met goede argumenten uit de literatuur.
<b>Respondent 7</b> Consultant  Publicaties in 2012 en 2016 Gericht op MKB organisatie en maturity model	-aantal vragen te moeilijk, -niet echt een relatie tussen dimensies en specifiek het MKB, komt niet duidelijk naar voren. -algemeen, vereenvoudigd de vragen verhoogt kans op toepassing.
<b>Respondent 8</b> Consultant  Publicaties in 2016 Gericht op MKB organisatie en maturity model	-enkele dimensies moeilijk beschreven

## Bijlage 13, Het FAM model, bij 1<sup>e</sup> iteratie toegepast.

### Gegevensbeschikbaarheid 1<sup>e</sup> iteratie

#### Inleiding

Maturity modellen zijn geïntroduceerd als referentiekaders voor het beheer van informatiesystemen (IS) in organisaties binnen verschillende industrieën. Bij MKB-organisaties blijft het gebruik van maturity modellen gericht op Data Analytics achter op grote organisaties. De toenemende hoeveelheid gegevens bij MKB-organisaties maakt het vermogen om deze te verwerken voor een verbeterde concurrentie positie steeds moeilijker. Aangenomen wordt dat zorgvuldig en aandachtig gebruik van Data Analytics bij MKB-organisaties gegevens kan omzetten in kennis die de bedrijfsresultaten en de operationele efficiëntie kan verbeteren. Belangrijke basis elementen bij het toepassen van Data Analytics zijn o.a. gegevensbeschikbaarheid en gegevenskwaliteit.

Dit onderzoek wil een kennis bijdrage leveren aan de knowledge base gericht op maturity modellen voor MKB organisaties. Tevens wil het MKB organisatie ondersteunen bij het toepassen van Data Analytics door een toepasbaar maturity model te ontwikkelen gericht op de belangrijke elementen gegevensbeschikbaarheid en gegevenskwaliteit.

De volgende onderzoeksvraag is geformuleerd:

**De hoofdvraag:** Hoe kunnen we een maturity model gericht op gegevensbeschikbaarheid en gegevenskwaliteit ontwikkelen ter ondersteuning van Data Analytics (DA) en is toepasbaar binnen een MKB-organisatie?

Om antwoord te krijgen op bovenstaande hoofdvraag zijn enkele deelvragen opgesteld. Een daarvan is:

**deelvraag:** Welke eisen kunnen aan gegevensbeschikbaarheid en gegevenskwaliteit worden gesteld voor een MKB-organisatie bij toepassing van Data Analytics ?

Om antwoord te vinden op deze vraag is er een model ontwikkeld voor gegevensbeschikbaarheid en voor gegevenskwaliteit. Beiden zijn met opzet gesplitst tijdens de iteratie om de omvang klein te houden. Hieronder het deel wat gericht is op gegevensbeschikbaarheid. Bedoeling is om via een enquête het uiteindelijke model bij een organisatie te toetsen.

#### Het concept maturity model, 'Gegevensbeschikbaarheid' gedeelte.

Diverse literatuur geeft aan dat de eisen welke gesteld kunnen worden aan gegevensbeschikbaarheid bij DA sterk context afhankelijk is. Dit onderzoek richt zich niet tot een specifieke context maar tot het toepassen van DA bij MKB organisatie in het algemeen. Wetende dat het bij een case studie een iets andere resultaten kan geven dan wanneer het gericht is op een specifieke context.

Het is opgezet vanuit de functie data analist of degene welke met deze taak is belast. Geen specifieke data-analyse voor een specifieke context, een core proces (bijv. inkoop, verkoop, warehouse, etc.) maar algehele status van data-analyse volwassenheid.

De dimensies zijn afkomstig uit bestaand onderzoek (Libes, X., 2015). Daarin worden een acht tal dimensies van gegevensbeschikbaarheid beschreven. De eerste dimensie binnen dat onderzoek beschrijft 'beperking van gegevensgebruik'. Verschillende literatuur (Bianchini et. al., 2019) (Baars, 2016) beschrijft het belang van zowel interne als externe data bij toepassing van Data-analyse. Om in het focus area maturity model de beschikbaarheid van data uit interne en externe bronnen te analyseren op volwassenheid is dit gesplitst (punt 1.1 en 1.2)

### Definitie

Gegevensbeschikbaarheid is gedefinieerd als: *de tijdigheid en betrouwbaarheid van toegang tot en gebruik van gegevens* (de graad van gemak voor gebruikers om gegevens en gerelateerde informatie te verkrijgen)

Het begrip gebruikers wordt in de context van Data Analytics niet alleen de data gebruikers gezien maar ook wat data-analyse (tools) zelf als gebruiker/handeling vereist om data te kunnen verwerken.

Het ontwikkelde FAM voor gegevensbeschikbaarheid:

1	Gegevensbeschikbaarheid	0	1	2	3	4	5	6	7	8	9	10	11	12
1.1	beperking in interne gegevensgebruik		A	B					C			D		
1.2	beperking in externe gegevensgebruik			A	B				C				D	
1.3	niet bestaande of ontbrekende gegevens					A	B			C			D	
1.4	timing en synchronisatie van gegevens			A			B				C			D
1.5	gegevensfrequentie			A		B	C				D	E		
1.6	Gegevensformat, standaarden en specificaties				A		B	C		D			E	
1.7	Onzekerheid, betrouwbaarheid en nauwkeurigheid				A		B				C			D
1.8	Toegang tot gegevens, opslag en verwerking					A	B		C			D		
1.9	Vorbereidingstijd van gegevens uit andere bronnen							A		B			C	
				DA niet of nauwelijks mogelijk		DA beperkt mogelijk			DA niet volledig mogelijk			DA volledig mogelijk		

Hierboven het focus area model, een matrix bestaande uit regels en kolommen. De negen focusgebieden zijn kenmerken van gegevensbeschikbaarheid gericht op Data Analytics met daarachter genummerde kolommen 0 t/m 12. Deze kolommen weerspiegelen de maturity levels (0 is geen, 12 is hoog). Elk focusgebied heeft unieke waarde welke varieert van drie tot vier letters wat de mate van tijdigheid en betrouwbaarheid van datatoegang en het gebruik vertegenwoordigd en gezamenlijk de algehele kwaliteit van DA beïnvloed.

#### De beoordeling van de verschillende maturity levels

In figuur 1 de opbouw van de weergave verder op. De dimensie en direct daaronder een korte uitleg. In de eerste kolom de maturity level en daarachter in de tweede kolom een status beschrijving. Direct onder de beschrijving de aanbevolen focus/aandachtgebied om uit te voeren om een volgend maturity level te behalen.

1.1 <Dimensie>	
< korte uitleg>.	
A: <maturity level>	<status beschrijving>
Focus:	<Focus / Aandachtsgebied om te ondersteunen in het behalen van de volgende maturity level>.

Figuur 1: voorbeeld van maturity level in de opzet hieronder

1.1 beperking in interne gegevensgebruik	
<i>De mate waarin de data analist tijdig en in voldoende mate over interne data kan beschikken.</i>	
A: Geen toegang tot data, DA niet mogelijk	Geen rechten of niet tijdig over benodigde data beschikken.
Focus:	<i>Aandacht gericht op inzicht in het belang van (hoeveelheid) data in relatie tot kwaliteit van DA. De toegang tot interne data voor DA functie/taak openstellen, vertrouwen krijgen in de data analist.</i>
B: Minimale toegang tot data, DA minimaal toegepast	Beperkte rechten, nauwelijks gebruik kunnen maken van beschikbare gegevens
Focus:	<i>Interne data toegang verbeteren door zoveel mogelijk interne data toegankelijk maken voor data analist. Bestanden in persoonlijke mappen op algemeen toegankelijke locaties plaatsen. Meer vertrouwen in de data analist krijgen en nemen. Zorg dat de analist overzicht krijgt van aanwezige interne data bronnen.</i>
C: Volledige toegang tot aantal datasets, DA mogelijk	Voldoende rechten, data tijdig beschikbaar, niet alle aanwezige bedrijfsdata beschikbaar voor Analyse.
Focus:	<i>Volledige medewerking en aanlevering van data aan de data analist bij verzoeken. Eventueel rechten uitbreiden op databases, zorg dat data analist kennis heeft van aanwezige tabellen en minimaal toegang heeft.</i>
D: Volledige toegang tot alle datasets	Alle aanwezige bedrijfsdata is tijdig beschikbaar en in voldoende mate toegankelijk
Focus:	<i>Blijft organisatie breed bewust van data hoeveelheid in relatie tot kwaliteit van data-analyse.</i>

1.2 beperking in externe gegevensgebruik
------------------------------------------



<i>In welke mate gebruikt de data analyst externe data of kan deze er over beschikken. Beperkingen kunnen zich voordoen om verschillende redenen zoals veiligheid en technisch (data verstrekker wil om (bedrijf- of productbescherming) niet alle data delen, real-time externe data vereist een continue toegang en dat vraagt om technische voorzieningen).</i>	
A: Geen toegang tot data, DA niet mogelijk	Geen gebruik maken van externe data, geen rechten of niet tijdig over benodigde data beschikken.
Focus:	<i>Vindt mogelijke externe data bronnen. Zorg waar mogelijk voor ingangen zodat bij een data behoefte deze ingang beschikbaar is.</i>
B: Minimale toegang tot data, DA minimaal toegepast	Beperkte gedeelde data ontvangen (of remote toegang), niet bekend of er andere mogelijke externe gegevens bronnen zijn
Focus:	<i>Verkrijg inzicht in zoveel mogelijke externe data bronnen. Tijdig opvragen externe data wanneer nodig, regel indien mogelijk een directe remote toegang.</i>
C: Volledige toegang tot aantal datasets, DA mogelijk	Volledige toegang tot externe data, geen automatische verversing / aanvulling met recente data.
Focus:	<i>Inventariseer of er via software een directe koppeling mogelijk is met eigen data-analyse pakketten.</i>
D: Volledige toegang tot alle datasets	Volledige toegang tot externe data, automatische verversing en of aanvulling met de meest recente data.
Focus:	<i>Ken de vereisten van de externe bronnen om toegang te behouden (abonnementskosten, periodiek vereiste wachtwoord verversing, regelmatig inloggen, etc).</i>

### 1.3 niet bestaande of ontbrekende gegevens

*Hoeveel data wordt er specifiek verzameld in de organisatie zodat dit voor DA kan worden gebruikt, zijn er bijvoorbeeld extra sensoren in de organisatie geplaatst zodat er extra data hoeveelheid verzameld wordt specifiek voor analyse. Indien er data verzameld wordt, weten we of de verzamelde data compleet is of kennen de datasets hiaten bijvoorbeeld het vergeten om een aantal formulieren in te vullen, storing aan een sensor waardoor deze geen gegevens kon doorgeven.*

A: Geen data verzameling specifiek voor DA	Er wordt niet specifiek voor data-analyse data verzamelt binnen de organisatie
Focus:	<i>Overweeg het gebruik van extra (data registratie) sensoren. Digitale sensoren of handmatige registratie gericht op plaatsen in het proces waar DA voor ingezet wordt.</i>
B: Ad hoc verzameling van data specifiek voor DA	Er wordt onregelmatig en/of beperkt voor analyse bedrijfseigen gegevens verzamelt. Er is geen zicht op volledigheid van de verzamelde data
Focus:	<i>Zorg voor regelmaat van sensor informatie. Weet wanneer sensor registratie uit is gebleven bij langdurige vastlegging. Vervang handmatige vastlegging van data naar geautomatiseerde.</i>
C: Voldoende data tijdig verzameld voor DA, betrouwbaarheid op compleetheid niet volledig	Er wordt regelmatig interne gegevens verzamelt specifiek voor analyse, we controleren via steekproef of deze volledig is.
Focus:	<i>Zorg voor data opslag capaciteit, maak via automatische controles zichtbaar dat er registraties ontbreken, bijvoorbeeld via logging van data opslag frequentie eventuele hiaten vinden.</i>

D: Voldoende data tijdig verzameld voor DA en maximale betrouwbaarheid.	Er wordt continue interne gegevens verzamelt specifiek voor analyse, we controleren de data structureel en weten altijd of deze volledig is
<i>Focus:</i>	<i>Controleer digitale sensoren met regelmaat en wees verzekerd van voldoende data verzameling capaciteit.</i>

#### 1.4 timing en synchronisatie van gegevens

*Bij data-analyse kan het moment (tijd) waarbij een waarde is vastgelegd belangrijk zijn. Vooral bij het gebruik van meerdere data bronnen is het van belang dat er zekerheid is zodat geregistreeerde tijden onderling vergelijkbaar zijn bijvoorbeeld chronologische volgorde van vastgelegde data.*

A: Tijdsanalyses onbetrouwbaar en of niet mogelijk	Er is geen inzicht in timing en synchronisatie m.b.t. data (datasets)
<i>Focus:</i>	<i>Analyseer de verzamelde data op het mogelijk vastleggen van tijdsregistraties. Verkrijg inzicht in waar timing en synchronisatie van belang is bij welke data-analyses.</i>
B: Tijdsanalyses onbetrouwbaar, niet structureel oplossing voor.	Het belang van timing en synchronisatie bij gegevens wordt onderkend, wordt indien mogelijk ad hoc (per analyse) opgelost.
<i>Focus:</i>	<i>Zorg dat bij vastlegging van data ook de tijden worden vastgelegd en dat belangrijke sensoren onderling geijkt zijn qua tijd/datum.</i>
C: Tijdsanalyses deels betrouwbaar, niet volledig structureel ingericht.	Aanwezigheid van aantal processen waarbij timing en sync standaard wordt meegenomen bij data verzameling zodat deze bij DA toegepast kunnen worden.
<i>Focus:</i>	<i>Zorg dat bij vastlegging van alle data ook de tijden worden vastgelegd waar nodig en deze onderling geijkt zijn zodat ongeacht de analyse er een vergelijking mogelijk is.</i>
D: Tijdsanalyses betrouwbaar, volledig structureel ingericht.	Bij alle processen, waarbij timing en synchronisatie van toepassing is, is de organisatie zich bewust van het belang en wordt dit opgeslagen voor gebruik bij DA.
<i>Focus:</i>	<i>Bewaak de timing en synchronisatie van de sensoren, zorg eventueel voor connectie naar een online time server.</i>

#### 1.5 Gegevensfrequentie

*De frequentie waarop interne data wordt verzameld / vastgelegd. (bijvoorbeeld bij sensoren op productiemachines). Meer beschikbare data kan leiden tot hogere nauwkeurigheid en betere inzichten, meer opslag betekend o.a. meer opslagruimte vereist en hogere kosten.*

A: Laag frequente verzameling.	Geen frequente opslag van data specifiek voor analyse, meer ad hoc / incidenteel, lage nauwkeurigheid voor DA
<i>Focus:</i>	<i>Weet welke analyse je wilt uitvoeren en welke informatie nodig is en wat er aan data verzamelt kan worden. Bepaal vooraf een verzamel periode.</i>
B: Ad hoc frequentie.	Voorafgaand aan Data Analytics zal specifieke data (mogelijk over een korte periode) worden verzameld om toegepast te kunnen worden. Voldoende data voor een gerichte DA, niet toepasbaar voor repeterende en onderling vergelijkbare analyses.
<i>Focus:</i>	<i>Zorg voor continue vastleggen van data op vaste meet punten in regelmatig te analyseren bedrijfsprocessen. Plaats digitale sensoren indien mogelijk, minder kans op menselijke fouten, hiaten.</i>

C: Frequente opslag met lage frequentie.	De frequentie van data opslag is elke dag gelijk. Sensoren staan bewust ingesteld op lage frequentie van gegevensopslag.
<i>Focus:</i>	<i>Verhoog de frequentie van data verzameling, denk aan opslag capaciteit. Meer data voor analyse levert meer betrouwbare analyses.</i>
D: Frequente opslag met hogere maar niet maximale frequentie.	De frequentie van data opslag is elke dag gelijk. Sensoren staan bewust ingesteld op gemiddelde frequentie van gegevensopslag.
<i>Focus:</i>	<i>Zet de frequentie van data opslag op maximaal. Plaats indien gewenst meer sensoren.</i>
E: Frequente opslag met maximale frequentie.	Er is continue opslag van data toepasbaar voor analyse, met bewust een hoge registratie frequentie zodat nauwkeurige analyse mogelijk.
<i>Focus:</i>	<i>Bewaakt de opslag capaciteit en het functioneren van de sensoren.</i>

## 1.6 Gegevensformat, standaarden en specificaties

*Elke organisaties kent ongelijksoortige gegevens en vaak opgeslagen in bestanden geformatteerd volgens verschillende normen. Machines welke data vastleggen en bijvoorbeeld bepaalde tijden waarbij de definitie van tijd niet duidelijk is kan problemen geven bij een analyse waarbij de tijd belangrijk is. (is het GMT of UTC tijd, is het 12 of 24 uren notatie). Kortom het normaliseren / standaardiseren van vast te leggen data kan van belang zijn. Bij standaardisatie kan gebruik gemaakt worden van Beschrijvende normen (descriptieve) indien technieken nieuw zijn en er nog geen norm is en Voorgeschreven normen (prescriptieve) indien er technische keuzes mogelijk zijn.*

A: Alleen analoog gegevens format aanwezig.	We gebruiken alleen papieren formulieren voor handmatige registraties. Voor DA gebruik eerst digitaliseren.
<i>Focus:</i>	<i>Handmatige data registraties waar mogelijk digitaal invoeren/registreren.</i>
B: Wisselende bestandsformaten aanwezig, geen standaardisatie.	Van de opgeslagen data is het format onbekend of kan het wisselen. Er is geen standaardisatie aanwezig, bij de uitvoering van DA is dit veelal een aandachtspunt.
<i>Focus:</i>	<i>Beperk de data invoer tot bekende formats en bij keuze beperk het aantal formats. Bijvoorbeeld alleen in een database of in Excel. Zorg bij aangeleverde externe data dat het format toegepast kan worden bij de analyse.</i>
C: Beperkte standaardisatie van data format aanwezig	Van bewaarde data is deels het format bekend en gestandaardiseerd. Bij de uitvoering van DA is dit een aandachtspunt, levert soms problemen op.
<i>Focus:</i>	<i>Zorg dat zoveel als mogelijk data voldoet aan een standaard format en geschikt is voor analyse. Standaardiseer de data formats van data welke frequent voor analyses gebruikt wordt.</i>
D: Meeste aanwezige data formats zijn volgens bekende normen.	Van meerdere opgeslagen data is format bekend en is de data uitwisselbaar volgens de standaard norm.
<i>Focus:</i>	<i>Weet hoe diverse data formaten verwerkt kunnen en vervolgens toegepast kunnen worden bij analyse. Standaardiseer zoveel mogelijk interne data opslag zodat deze altijd geschikt is voor toekomstige analyse. Gebruik analyse tools welke overweg kunnen met verschillende data formats.</i>
E: Alle aanwezige data formats zijn volgens bekende normen.	Alle data formats zijn bekend en volledig uitwisselbaar volgens voorgeschreven standaardisatie. Waar nodig helpt de DA tooling bij de verschillende formats.

	Focus:	Bewaak het formaat waarin data wordt aangeleverd voor analyse. Weet de data format mogelijkheden van de gebruikte analyse tools.
--	--------	----------------------------------------------------------------------------------------------------------------------------------

### 1.7 Onzekerheid, betrouwbaarheid en nauwkeurigheid

*Bedrijfsgegevens kunnen kenmerken hebben welke van invloed zijn op Data Analytics zoals onzekerheid, betrouwbaarheid en nauwkeurigheid. Het is belangrijk dat er bij de te verwerken data een begeleidend schrijven is wat de data beschrijft (metadata). Dit kan o.a. door bij analyse een beschrijving te verstrekken via gekwantificeerde onzekerheid en betrouwbaarheid over de data.*

A: Geen zekerheid over data betrouwbaar- en nauwkeurigheid.		We controleren niet bewust de betrouwbaarheid / nauwkeurigheid van de data voor DA gebruik. We gaan er vanuit dat de gebruikte data betrouwbaar is.
	Focus:	<i>Bij herhalende analyses verschaft inzicht in mogelijke punten binnen de data waar steekproef controle mogelijk is. Stel voorafgaand aan DA vast hoe betrouwbaar de data kan zijn (bron controle, is data actueel/bijgewerkte).</i>
B: Enige mate van zekerheid over data betrouwbaar- en nauwkeurigheid.		We controleren bewust de betrouwbaarheid / nauwkeurigheid van data bij DA. We zijn ons bewust van het belang bij DA. We hebben geen extra beschrijving bij de data waarin dit beschreven is.
	Focus:	<i>Controleer vooraf of er metadata beschikbaar is bij de dataset en neem dit door.</i>
C: Er is kortstondig zekerheid over data betrouwbaar- en nauwkeurigheid bij data-analyse.		Voorafgaand aan een analyse doen we een audit op data betrouwbaarheid / nauwkeurigheid. Eventueel aanwezige metadata wordt doorgenomen.
	Focus:	<i>Creëer bewustwording binnen de organisatie gericht op data betrouwbaarheid en nauwkeurigheid. Creëer bij data vastlegging de mogelijkheid voor mensen / machines om opmerkingen of Tag's te plaatsen (in metadata) waar twijfels bestaand over betrouwbaarheid of nauwkeurigheid.</i>
D: Er is continue zekerheid over data betrouwbaar- en nauwkeurigheid bij data-analyse door structurele bewaking		Gebruikte data is betrouwbaar en nauwkeurig bij toepassen van DA. Waar mogelijk lezen we de metagegevens en verwerken we dit bij de resultaatgegevens. Regelmatig doen we audit/controle steekproef gericht op (RAW) data betrouwbaarheid en nauwkeurigheid
	Focus:	<i>Bewaak de audits en maak resultaten kenbaar binnen de organisatie.</i>

### 1.8 Toegang tot gegevens, opslag en verwerking

*Gegevenstoegang heeft invloed op de gegevensanalyse. Qua toegang is de vraag of we direct op de productie data zitten of dat voor DA met een kopie gewerkt wordt. Bij dit laatste moeten we vervolgens rekening houden met het feit dat data onbewust aangepast kan zijn zodat het als kopie opgeslagen kan worden denk een grote dataset waar alleen maar een klein deel nodig is. Data kan in verschillende type databases opgeslagen zijn. Elke heeft zijn voor- nadelen bij verwerking in DA. Dit speelt vooral een rol bij ondernemingen met meerdere (typen) databases. Tabellen samenvoegen binnen een database is relatief eenvoudig. Tabellen koppelen uit verschillende databases vraagt om specifieke kennis.*

A: DA kan alleen gebruik maken van data kopie bestanden	Toegang tot data voor DA gaat via data kopieën. Of deze mogelijk is aangepast weten we verder niks van / doen we niks mee.
<i>Focus:</i>	<i>Weet welke data kopieën mogelijk aangepast zijn, en doe daarbij een controle met de originele data. Zorg waar mogelijk dat toekomstige kopieën structureel geen data wijzigingen bevatten, ken in ieder geval de probleem kopieën. (denk bijv. aan export van berekende velden)</i>
B: DA kan gebruik maken van kopie en of live data	Toegang tot data voor DA gaat via data kopieën of direct naar een database of databestanden. Van de kopie data weten we dat deze mogelijk is aangepast.
<i>Focus:</i>	<i>Zorg voor een structureel mechanisme (Checksums) wat de kopie data controleert met het origineel.</i>
C: DA gebruikt kopie en of live data, tabellen samenvoegen is geen probleem	Van de kopie data weten we dat er geen data is aangepast, eventuele tabellen samenvoegen voor DA kan waar nodig.
<i>Focus:</i>	<i>Vergroot de kennis van databases, vooral het koppelen van verschillende databases en tabellen. Gebruik eventueel software welke hierbij ondersteund.</i>
D: DA maakt gebruik van kopie en of live data, het maakt gebruik van tabellen samenvoegen en of koppelt met andere verschillende databases.	Er is een directe koppeling naar de data bron mogelijk, DA maakt gebruik van kopie en of live data, het maakt gebruik van tabellen samenvoegen en of koppelt met andere verschillende databases. (bijv. gebruik van data warehouse)
<i>Focus:</i>	<i>Zorg bij kopie data, waarop belangrijke data-analyses zijn gedaan dat voor deze kopie data een back-up beleid is en eventuele analyses opnieuw gedaan kunnen worden. (aangenomen dat voor live data als een back-up policy bestaat en hierop gecontroleerd wordt)</i>

### 1.9 Voorbereidingstijd van gegevens uit andere bronnen

*Het verkrijgen van data uit andere (intern / extern) niet standaard bronnen vereist voorbereiding voor verwerking. Voor verwerking en interpretatie dient de ontvangen data geanalyseerd te worden voorafgaand aan verdere verwerking binnen DA en eventuele metadata gegevens doorgenomen te worden.*

A: DA wordt uitgevoerd zonder voorbereiding.	Bij ontvangen data uit andere bronnen bereiden we nauwelijks iets voor, beginnen vrijwel direct met analyse. We bezitten niet voldoende de kennis om eventueel data te prepareren voor DA / we vertrouwen erop dat de datasets goed zijn / er voor interpretatie van de data geen extra informatie nodig is / er voldoende capaciteit voor analyse is (bewerking en opslag).
<i>Focus:</i>	<i>Plan structureel vooraf tijd in gericht op controle van ontvangen data, controleer op onregelmatigheden (ontbreken van bijv. datavelden, is de data omvang aannemelijk), controleer of verwerkingscapaciteit voldoende is. Weet van regelmatig benodigde en ontvangen data wat er aan data preparatie nodig is voor DA.</i>
B: Afhankelijk van de databron is er een voorbereiding.	Bij ontvangen data uit andere bronnen bereiden we afhankelijk van de bron iets voor, en beginnen vervolgens met analyse. We vertrouwen er op dat de uitgevoerde

	voorbereiding goed zijn, er voldoende capaciteit is voor de analyse.
<i>Focus:</i>	<i>Weet structureel of data ook metadata bevat, probeer bij belangrijke data dit structureel te verkrijgen en desnoods structureel vast te leggen bij de interne data verzameling. Zorg voor het kunnen toetsen van DA uitkomsten, bekijk of er bijzondere standaardafwijkingen zijn. Verwerkt / registreer metadata informatie in de analyse.</i>
C: ongeacht de databron is er een voorbereiding.	Nieuwe DA's starten we altijd met specifieke vastgestelde voorbereidingshandelingen. We verzekeren er ons van dat er voldoende capaciteit voor bewerking en opslag hebben.
<i>Focus:</i>	<i>Er is een vastgesteld proces van data verwerking bij DA waarbij een vast aantal data beoordelingsacties voor en na analyse aanwezig zijn.</i>

# Gegevenskwaliteit 1<sup>e</sup> iteratie

## Inleiding

Maturity modellen zijn geïntroduceerd als referentiekaders voor het beheer van informatiesystemen (IS) in organisaties binnen verschillende industrieën. Bij MKB-organisaties blijft het gebruik van maturity modellen gericht op Data Analytics achter op grote organisaties. De toenemende hoeveelheid gegevens bij MKB-organisaties maakt het vermogen om deze te verwerken voor een verbeterde concurrentie positie steeds moeilijker. Aangenomen wordt dat zorgvuldig en aandachtig gebruik van Data Analytics bij MKB-organisaties gegevens kan omzetten in kennis die de bedrijfsresultaten en de operationele efficiëntie kan verbeteren. Belangrijke basis elementen bij het toepassen van Data Analytics zijn o.a. gegevensbeschikbaarheid en gegevenskwaliteit.

Dit onderzoek wil een kennis bijdrage leveren aan de knowledge base gericht op maturity modellen voor MKB organisaties. Tevens wil het MKB organisatie ondersteunen bij het toepassen van Data Analytics door een toepasbaar maturity model te ontwikkelen gericht op de belangrijke elementen gegevensbeschikbaarheid en gegevenskwaliteit.

De volgende onderzoeksvraag is geformuleerd:

**De hoofdvraag:** Hoe kunnen we een maturity model gericht op gegevensbeschikbaarheid en gegevenskwaliteit ontwikkelen ter ondersteuning van Data Analytics (DA) en is toepasbaar binnen een MKB-organisatie?

Om antwoord te krijgen op bovenstaande hoofdvraag zijn enkele deelvragen opgesteld. Een daarvan is:

**deelvraag:** Welke eisen kunnen aan gegevensbeschikbaarheid en gegevenskwaliteit worden gesteld voor een MKB-organisatie bij toepassing van Data Analytics ?

Om antwoord te vinden op deze vraag is er een model ontwikkeld voor gegevensbeschikbaarheid en voor gegevenskwaliteit. Beiden zijn met opzet gesplitst tijdens de iteratie om de omvang klein te houden. Hieronder het deel wat gericht is op gegevenskwaliteit. Bedoeling is om via een enquête het uiteindelijke model bij een organisatie te toetsen.

## Het concept maturity model, 'Gegevenskwaliteit' gedeelte.

Diverse literatuur geeft aan dat de eisen welke gesteld kunnen worden aan gegevenskwaliteit bij DA sterk context afhankelijk is. Dit onderzoek richt zich niet tot een specifieke context maar tot het toepassen van DA bij MKB organisatie in het algemeen. Wetende dat het bij een case studie mogelijk iets andere resultaten geeft dan wanneer het gericht is op een specifieke analyse.

Het is opgezet vanuit de data analyst rol of degene welke met deze taak is belast. Geen specifieke data-analyse voor een core proces (bijv. inkoop, verkoop, warehouse, etc) maar algehele status van data-analyse volwassenheid.

De dimensies zijn afkomstig uit een onderzoek (Merino, J., Ismael, C., Bibiano, R., Manuel, S. en Mario, P., 2015) wat is gebaseerd op de klassieke indeling van Strong (1997) het 'Data Kwaliteit-in-gebruik' model.

Een vijftien tal dimensies van gegevenskwaliteit beschreven en onderverdeeld naar contextuele geschiktheid, tijdelijke geschiktheid en operationele geschiktheid. Contextuele geschiktheid verwijst naar "het vermogen van datasets om te worden gebruikt binnen hetzelfde interessegebied van de analyse, onafhankelijk van welk formaat dan ook (bijv. gestructureerd versus ongestructureerd), elke grootte of de snelheid van instroom". Tijdelijke geschiktheid verwijst naar "data binnen een passend tijdvak voor de analyse" (bijv. vergelijkbare leeftijd of gelijktijdige gegevens). Dat de gegevens verwijzen naar een soortgelijk periode. De laatste operationele geschiktheid, betreft de mate waarin gegevens kunnen worden verwerkt in de beoogde analyse en geschikte methode zonder dat er gegevens (deels) buiten de analyse vallen. Dit betekent dat er voldoende en geschikte middelen zijn om de analyse uit te voeren (bijv. vergelijkbare gegevenstypen).

### Definitie

Gegevenskwaliteit is in dit onderzoek gedefinieerd als: *de mate waarin de kwaliteit voldoet aan, of deze de kwaliteitseisen overtreft, die gebruikers ervan verwachten.*

Het begrip gebruikers wordt in de context van Data Analytics niet alleen de data gebruikers gezien maar ook wat data-analyse (tools) zelf als gebruiker/handeling vereist om data te kunnen verwerken.

### De Maturity levels onderling

Tussen de verschillende dimensies levert bestaande literatuur geen antwoord op of er een eventueel onderling belangen verschil is of dat er mogelijk afhankelijkheden zijn. Het ontbreken van onderlinge belangen (verhoudingen) tussen de dimensies levert voor zover nu inzichtelijk, geen extra toegevoegde waarde. Daarom zijn alle A levels onder één kolom geplaatst, B onder een volgende geplaatst, etc.

Hieronder het focus area model, een matrix bestaande uit regels en kolommen. Drie focusgebieden in de eerste kolom met elk een aantal dimensies (kenmerken) van gegevenskwaliteit en gericht op Data Analytics .





2.1 Nauwkeurigheid - Accuracy	
<i>In hoeverre de gegevens kenmerken hebben die de werkelijke waarde van het beoogde concept of gebeurtenis in een specifieke gebruikscontext correct weergeven nodig voor data-analyse.</i>	
A: Nauwkeurigheid niet of nauwelijks toegepast	Geen idee over de nauwkeurigheid van de data voor aanvang van data-analyse
Focus:	<i>Creëer bewustwording van data nauwkeurigheid, erken het belang ervan.</i>
B: Nauwkeurigheid van data wordt erkend, vertrouwen erop data dit in voldoende mate bezit.	We zijn ons bewust van het belang van data nauwkeurigheid, verder doen we geen specifieke controle bij het uitvoeren van data-analyse
Focus:	<i>Leg bij de belangrijkste data vast hoe deze op nauwkeurigheid gecontroleerd kan worden en bij herhaalde analyses op dezelfde wijze kan worden uitgevoerd.</i>
C: Nauwkeurigheid van data wordt erkend, aanwezigheid een aantal middelen waardoor steekproef controle kan worden uitgevoerd over een deel van de aangeleverde data.	We zijn ons bewust van het belang van data nauwkeurigheid, we controleren waar mogelijk data vooraf (steekproef) of deze de waardes bevat welke afkomstig zijn van de verwachte data bron, of het aangeleverde data format geen belemmering oplevert en bij herhaalde analyse dezelfde data velden weer aanwezig zijn zodat vergelijking mogelijk is
Focus:	<i>Breidt de controle mogelijkheid uit naar zoveel als mogelijk.</i>
D: Aanwezigheid van controle middelen op data nauwkeurigheid. Dit wordt structureel toegepast.	Gebruikte data is nauwkeurig, we hebben de zekerheid dat de data afkomstig is van de verwachte data bron, het data format is niet van invloed op de analyse. Structureel en voorafgaand aan data-analyse wordt hier op gecontroleerd
Focus:	<i>Vastgelegen hoe nauwkeurigheidscontroles structureel uitgevoerd worden per analyse.</i>
E: organisatie is ervaren in het doen van audits gericht op data nauwkeurigheid.	Regelmatig doen we audit/controle steekproef gericht op nauwkeurigheid van data ten behoeve van data-analyse.
Focus:	<i>Frequentie van audits vastleggen.</i>

2.2 Volledigheid – Completeness	
<i>De hoeveelheid gebruikte gegevens passend en relevant in het kader van de analyse.</i>	
A: Geen inzicht in data volledigheid	Er is geen controle of de gebruikte data voor data-analyse volledig is, we doen geen controle.
Focus:	<i>Creëer bewustwording van data volledigheid, erken het belang ervan.</i>
B: Beperkt inzicht of data hoeveelheid past bij de analyse.	We zijn ons bewust van het belang van data volledigheid bij data-analyse, we accepteren eventuele afwijkingen na analyse.
Focus:	<i>Leg bij de belangrijkste data vast hoe deze op volledigheid gecontroleerd kan worden en bij herhaalde analyses op dezelfde wijze kan worden uitgevoerd.</i>
C: Erkenning van data volledigheid, controle wordt waar mogelijk toegepast.	We zijn ons bewust van het belang van data volledigheid bij data-analyse, we controleren voorafgaand aan data-analyse of de hoeveelheid data passend is voor de concrete analyse.
Focus:	<i>Bepaal en leg vast hoeveel data vereist is voor de belangrijkste analyses om voldoende kwaliteit te krijgen. Zorg voor overzicht van de verschillende formats waarin data is aangeleverd.</i>

D: Volledigheidscontroles worden waar mogelijk toegepast en uitgebreid.	Concrete acties lopen om data volledigheid waar nodig te verbeteren. Van eventuele data format problemen zijn we ons bewust.
<i>Focus:</i>	<i>Bepaal en leg vast hoeveel data vereist is voor de meeste analyses om voldoende kwaliteit te krijgen en controleer voorafgaand aan analyse of de gestelde hoeveelheid daadwerkelijk aanwezig is. Weet van veel voorkomende formats hoe deze waar nodig aangepast moeten worden.</i>
E: Volledigheidscontroles vinden vooraf gaan aan de analyse plaats bij belangrijke analyses.	We zijn ons bewust van het belang van data volledigheid bij data-analyse. We trachten naar inzicht altijd de data volledig te hebben alvorens de analyseren. Aangeleverde dataset formats worden vooraf aangepast waar nodig zodat er geen beperking is bij analyse. Indien mogelijk wordt datasets aangevuld zodat deze volledig zijn of wordt dit na analyse vermeld in het begeleidende schrijven.
<i>Focus:</i>	<i>Leg vast bij elke analyse waarop binnen de dataset gecontroleerd dient te worden om de volledigheid te toetsen.</i>
F: Volledigheidscontroles vinden structureel plaats	Voorafgaand op data-analyse doen we audit/controle steekproef gericht op volledigheid van verzamelde data
<i>Focus:</i>	<i>Bewaak de periodieke audits en de kwaliteit ervan. Leg de uitvoering en de bevindingen vast.</i>

### 2.3 Consistentie – Consistency

*De gegevens mogen geen hiaten bevatten in het tijdspad van analyse (tijdconsistentie), geen data duplicaten en vrij van tegenstrijdigheden. Hebben meet sensoren bijvoorbeeld continue goed gefunctioneerd en data constant aangeleverd (bijv. vergelijkbare gegevens voor vergelijkbare entiteiten).*

A: inconsistente data gebruik mogelijk aanwezig.	We weten niet of data voor Data-analyse tijdspad hiaten bevat en of er dubbele waardes in zitten. We doen geen controle.
<i>Focus:</i>	<i>Leg vast hoe data deduplicatie kan worden toegepast bij de analyses. Bij tijdsanalyses controleer via steekproef of op verwachte momenten ook een registratie is gedaan of zijn er blanco tijdsperiodes (achtereenvolgende momenten waarop data zou moeten zijn vastgelegd).</i>
B: enige mate van data inconsistentie wordt geaccepteerd.	We zijn ons bewust dat data mogelijk iets van een hiaat bevat, maar we denken dat dit bij de analyses welke we regelmatig uitvoeren de uitkomst beperkt beïnvloed.
<i>Focus:</i>	<i>Zorg voor voldoende kennis en of technieken om hiaten in data consistentie te controleren, gebruik eventueel metadata om inconsistentie bij data verzameling vast te leggen.</i>
C: structurele controle op data consistentie, dit wordt in meta data vastgelegd.	Data wordt vooraf gecontroleerd op mogelijke in consistentie en wordt dit ter kennis meegenomen.
<i>Focus:</i>	<i>Zorg voor voldoende kennis om data hiaten indien mogelijk aan te vullen, weet waar ontbrekende data gehaald kan worden. Als aanvulling niet kan, weet hoe de analyse afgerond kan worden met toevoeging van de juiste opmerking.</i>
D: structurele controle op data consistentie en herstel indien mogelijk.	Data wordt vooraf gecontroleerd op mogelijke in consistentie en wordt dit waar mogelijk hersteld alvorens met data-analyse te beginnen.

	<i>Focus:</i> Zorg dat kennis en techniek aanwezig blijft om hiaten in data te detecteren.
--	--------------------------------------------------------------------------------------------

#### 2.4 Geloofwaardigheid - Credibility (Believability)

*De mate waarin gegevens kenmerken hebben die in een specifieke gebruikscontext als waar en geloofwaardig (authentiek) kunnen worden beschouwd en de verzameling op een juiste wijze heeft plaatsgevonden.*

A: Geen controle op data geloofwaardigheid.	Vooraf aan DA controleren we niet (via steekproef) of data geloofwaardig is.
<i>Focus:</i>	<i>Leg vast hoe bij een aantal (herhalende) analyses op geloofwaardigheid gecontroleerd kan worden. Is bijvoorbeeld de data gegarandeerd van de verwachte sensor of machine, is de data gelijk aan hetgeen verwacht kan worden kwantitatief, format, type, etc.</i>
B: Enige mate van data geloofwaardigheidscontrole aanwezig.	Vooraf aan DA soms een steekproef of data geloofwaardig is, dat deze afkomstig is van de verwachte data bron, dat de omvang in lijn is met de verwachte omvang. Methode van verzamelen wordt niet gecontroleerd.
<i>Focus:</i>	<i>Zorg voor databron vermelding bij data verzameling, waar mogelijk een controle van tijd registraties periodiek uitvoeren. Vergaar inzicht in mogelijke data verzamel methode en welke toegepast worden. Creëer methode om herhaalde data geloofwaardigheid controle te kunnen uitvoeren.</i>
C: Structurele data geloofwaardigheidscontroles worden uitgevoerd.	Vooraf aan DA altijd een steekproef of data geloofwaardig is, of de te verwachten data ook in de dataset aanwezig is. Indien nodig vindt er controle plaats op de wijze van verzamelen.
<i>Focus:</i>	<i>Waar mogelijk analyse resultaten controleren op bijzondere uitkomsten en zo nodig terugkoppelen naar mogelijke oorzaken in de verzamelde data.</i>
D: Regelmatige controle van DA resultaten op geloofwaardigheid van de uitkomst.	Na analyse worden de resultaten regelmatig gecontroleerd, om mogelijke ongeloofwaardige data uitkomsten (fouten in de ruwe data) achteraf te achterhalen. Methode van verzamelen is bekend en bij twijfel doen we een controle.
<i>Focus:</i>	<i>Verkrijg kennis/ervaring in het controleren van analyse resultaten op geloofwaardigheid en het herleiden naar verzamelde data. Wijzig indien nodig de methode van verzamelen.</i>
E: Structurele geloofwaardigheidscontroles voor en na analyse.	Analyse resultaten worden altijd gecontroleerd op geloofwaardigheid, om mogelijke ongeloofwaardige data uitkomsten (fouten in de ruwe data) te achterhalen. De methode van verzamelen is en wordt met regelmaat beoordeeld.
<i>Focus:</i>	<i>Bewaken van data verzamelmethodes en leg uitgevoerde audits vast.</i>

#### 2.5 Actueel - Currentness

*De mate waarin gegevens kenmerken hebben die van de juiste leeftijd zijn binnen een specifieke gebruikscontext.*

A: geen controle op actualiteit van gebruikte data	We gebruiken de data welke nodig is, er is geen controle of deze op moment van DA ook actueel is.
----------------------------------------------------	---------------------------------------------------------------------------------------------------

	<i>Focus:</i>	<i>Bekijk de aangeleverde data vooraf en bekijk of het datum – tijd informatie bezit wat beoordeeld kan worden. Weet vooraf aan de analyse welke periode van belang is.</i>
B: geen controle op actualiteit van gebruikte data, na analyse soms controle op actualiteit.		We gebruiken de data welke nodig is, er is geen controle of deze op moment van DA ook actueel is. Mogelijk dat de resultaten na analyse vragen oproepen over de actualiteit.
	<i>Focus:</i>	<i>Weet bij welke herhaalde analyses datum tijd van belang is, zorg voor kennis hoe dit gecontroleerd kan worden.</i>
C: intuïtieve controle van data op actualiteit.		We weten dat actuele data van belang is, we controleren dit incidenteel voorafgaand aan de analyse.
	<i>Focus:</i>	<i>Zorg voor (automatische) structurele controles van veelgebruikte data op datum- en tijdwaardes zodat bij gebruik voor analyse de actualiteit kenbaar is.</i>
D: structurele controle van data op actualiteit.		We weten vooraf dat de data actueel is, dat de ruwe data alleen de benodigde data van een specifiek tijdspad bevat.
	<i>Focus:</i>	<i>Bewaak de structurele controles op data.</i>

## 2.6 Toegankelijkheid - Accesibility

*De mate waarin gegevens toegankelijk zijn in een specifieke gebruikscontext, met name door mensen die ondersteunende technologie of speciale configuratie nodig hebben vanwege een toegangsbeperking.*

A: voor DA een te beperkte data toegang.		Er is (deels) geen toegang tot de gegevens.
	<i>Focus:</i>	<i>Zorg voor minimaal toegang tot een kopie van data bestanden of database. Pas desnoods de rechten aan zodat er in ieder geval deels toegang is tot data voor analyse.</i>
B: Indirecte data toegang via kopie live data.		Er is indirect toegang tot gegevens middels een data kopie en aangepaste rechten, real-time data-analyse is niet mogelijk daardoor.
	<i>Focus:</i>	<i>Zorg voor een datakopie welke zich kan synchroniseren met de live data zodat actuele data-analyses (op live data) mogelijk is.</i>
C: Directe en indirecte toegang tot data.		Er is direct toegang tot kopie gegevens waarbij real-time data-analyse mogelijk is (kopie data bron synchroniseert met data bron).
	<i>Focus:</i>	<i>Zorg voor een directe data toegang en dat de rechten voldoende zijn zodat data-analyses op live data mogelijk is.</i>
D: Directe toegang tot live data.		Er is een directe koppeling naar de data bron, dit is eventueel via aangepaste rechten mogelijk gemaakt.
	<i>Focus:</i>	<i>Zorg dat data analist naast bestaande ook bij nieuwe data bronnen voldoende rechten krijgen bij inrichting.</i>

## 2.7 Nakomen - Compliance

*De mate waarin gegevens kenmerken hebben die voldoen aan geldende normen, conventies of voorschriften en soortgelijke regels met betrekking tot gegevenskwaliteit in een specifieke gebruikscontext.*

A: onbekend of niet compliance		Er zijn mogelijk branchevoorschriften en of wettelijke regels voor onze data opgesteld, we weten het niet.
	<i>Focus:</i>	<i>Achterhaal of er branchevoorschriften en of wettelijke regels bestaan voor data verwerking.</i>

B: data compliance meer verankerd in de informele organisatie	Er zijn branchevoorschriften en of wettelijke regels voor onze data opgesteld, we voldoen er niet aan
<i>Focus:</i>	Verwerk de belangrijkste branchevoorschriften en of wettelijke regels bij de data verwerking
C: data compliance meer verankerd in de formele organisatie	Er zijn branchevoorschriften en of wettelijke regels voor onze data opgesteld, we voldoen er deels aan
<i>Focus:</i>	Verwerk de branchevoorschriften en of wettelijke regels bij de data verwerking
D: data is in compliance	Er zijn branchevoorschriften en of wettelijke regels voor onze data opgesteld, we voldoen er volledig aan
<i>Focus:</i>	Zie toe op het hanteren van de branchevoorschriften en of wettelijke regels bij de data verwerking.

## 2.8 Vertrouwelijkheid - Confidentiality

*De mate waarin gegevens kenmerken hebben die ervoor zorgen dat ze alleen toegankelijk en interpreteerbaar zijn in een specifieke gebruikscontext. Vertrouwelijkheid is een aspect van informatiebeveiliging (samen met beschikbaarheid, integriteit).*

A: geen vertrouwelijkheids beperkingen instelbaar	Meerdere data is door verschillende mensen benaderbaar, niet alleen de data analist, het is begrijpelijk voor veel mensen.
<i>Focus:</i>	Probeer data toegang af te schermen van hen welke qua functie geen toegangsbelang hebben
B: Beperkte vertrouwelijkheids beperkingen instelbaar	Toegangsrechten aanwezig zodat niet iedereen zomaar toegang heeft tot alle data, zij die toegang hebben begrijpen de data in zijn context.
<i>Focus:</i>	Vergroot het vertrouwen in de data analist, beperkt zoveel mogelijk de toegang tot de data door anderen.
C: Volledige vertrouwelijkheids beperkingen instelbaar	Alleen de data analist heeft toegang en begrijpt de data.
<i>Focus:</i>	Maximaliseer het vertrouwen in de data analist.

## 2.9 Efficiëntie - Efficiency

*De mate waarin gegevens kenmerken hebben die kunnen worden verwerkt en de verwachte prestatieniveaus bieden met behulp van de juiste hoeveelheden en typen resources in een specifieke gebruikscontext. (bijvoorbeeld een pdf eenvoudig om te zetten naar een werkbaar / doorzoekbaar bestand, datacompressie waardoor 'lege' ruimtes verwijderd worden in een database). Zeker van belang bij grote hoeveelheden data voor analyse. Denk ook aan data reductie mogelijkheden (data compressie\* / data deduplicatie\*\*), indexeren (PDF bestanden)*

A: Geen data efficiëntie	Data bestanden kunnen we niet wijzigen zodat deze mogelijk minder efficiënt toepasbaar is voor DA.
<i>Focus:</i>	Zorg dat er kennis is hoe belangrijke datasets aangepast kunnen worden zodat verwerking ervan voor analyse verbeterd.
B: Beperkte data efficiëntie	Sommige datasets kunnen we wijzigen zodat deze efficiënt toepasbaar zijn voor DA gebruik.
<i>Focus:</i>	Zorg dat alle data indien nodig aangepast kan worden zodat verwerking ervan voor analyse mogelijk is.

C: efficiëntie	Vooraf aan data-analyse kunnen we data bestanden wijzigen zodat deze beter toe te passen zijn bij DA,
<i>Focus:</i>	<i>Zorg dat alle datasets indien nodig aangepast kan worden zodat er een efficiëntere verwerking mogelijk is. Wellicht kan analyse software uitkomst bieden.</i>
D: Maximale efficiëntie	Vooraf aan data-analyse wordt altijd (evt. automatisch) de te gebruiken datasets en bestanden waar mogelijk gewijzigd zodat deze beter toe te passen zijn bij DA. We zijn in staat om alle data nodig voor analyse efficiënt te gebruiken.
<i>Focus:</i>	<i>Beoordeel regelmatig of er efficiëntie behaald kan worden.</i>

**\*Datacompressie**

Met deze techniek wordt de informatie gecodeerd met behulp van minder databits. Compressie-algoritmen kunnen verlieslatend (informatie gaat verloren, resolutie data neemt af) en verliesvrij zijn (alles wordt behouden door verwijderen van statistische redundantie)

**\*\*Data deduplicatie**

Hiermee elimineer je dubbele kopieën van data binnen een opslagvolume of over het gehele opslagsysteem. Deduplicatie maakt gebruik van patroonherkenning om overbodige data te identificeren. Deze worden vervolgens vervangen door verwijzingen naar een enkele opgeslagen kopie.

2.10 Precisie - Precision	
<i>De mate waarin gegevens kenmerken bevatten die exact zijn of die discriminatie opleveren in een specifieke gebruikscontext.</i>	
A: Onbekend met data precisie	We weten niet hoe of kijken niet naar data of deze afhankelijk van de analyse verschillende waardes kan opleveren.
<i>Focus:</i>	<i>Verdieping in het toepassen van verschillende analyse methodes</i>
B: Enigszins bekend met data precisie.	We weten dat data bij analyse verschillende waardes kan leveren, we doen er verder niks mee waarschijnlijk niet relevant voor onze analyses.
<i>Focus:</i>	<i>Verdieping in toepassen van verschillende analyse methodes en kennis van mogelijk verschillende uitkomsten.</i>
C: Bewust bezig met data precisie.	Bij een aantal analyses (herhalende) kijken we bewust naar data en of deze afhankelijk van type analyse verschillende waardes kan leveren.
<i>Focus:</i>	<i>Probeer verschillende analyse methodes op dezelfde data en beoordeel de uitkomsten op verschillen. Zoek per context naar goede (beste) analyse.</i>
D: Data precisie is een vaste handeling bij analyse.	We kijken structureel naar data eigenschappen en weten dat deze afhankelijk van de analyse verschillende waardes kunnen leveren.
<i>Focus:</i>	<i>Zorg dat kennis van analyse methodes behouden en actueel blijft.</i>

2.11 Traceerbaarheid- Traceability	
<i>De mate waarin gegevens kenmerken hebben die een audit trail (controlespoor) bieden van toegang tot de gegevens en van eventuele wijzigingen in de gegevens in een specifieke gebruikscontext.</i>	
A: Traceerbaarheid van toegang en wijziging niet mogelijk en of onbekend	Van data bestanden of databases weten we niet of deze naderhand nog aangepast zijn, kunnen we niet controleren, er is geen mutatie registratie.

	<i>Focus:</i>	<i>Weet hoe en waar je bij verschillende bestandtypes de meta data kan uitlezen. Achterhaal of op databases een traceerbaarheid mogelijk is.</i>
B: Traceerbaarheid beperkt mogelijk		Op gedeeltes van de aanwezige data kunnen we controleren op mutaties naderhand.
	<i>Focus:</i>	<i>Schakel aanwezige traceerbaarheids mogelijkheden op databases in, controleer altijd op aanwezige metadata in bestanden.</i>
C: Traceerbaarheid volledig mogelijk		We controleren daar waar mogelijk altijd op mogelijke naderhand nog gedane mutaties in data. Soms doen we dit bij DA vooraf en soms na het toepassing van DA.
	<i>Focus:</i>	<i>We hebben vastgelegd wanneer we bij DA in het begeleidend schrijven informatie vastleggen over gemuteerde data wanneer dit mogelijk van belang is bij interpretatie van het resultaat.</i>

## 2.12 Begrijpelijkheid - Understandability

*De mate waarin gegevens kenmerken hebben die het mogelijk maken om deze door gebruikers te lezen en te interpreteren, en worden uitgedrukt in geschikte talen, symbolen en eenheden in een specifieke gebruikscontext. Sommige informatie beschikt over metadata, een begeleidend schrijven.*

A: Onbekend of aanwezige data extra informatie bevat.		Bij DA verwerken we data maar hebben geen kennis van deze (ruwe) data. De verkregen informatie wordt door anderen geïnterpreteerd voor verdere kennis verwerking.
	<i>Focus:</i>	<i>Vergaar inzicht in het belang van de ondersteunende functie van metadata en achterhaal wat er reeds van aanwezig is.</i>
B: Bewust van data begrijpelijkheid. Incidenteel wordt dit toegepast ter ondersteuning van de analyse resultaten.		Bij DA verwerken we data en kijken bij sommige data naar de bijbehorende metadata.
	<i>Focus:</i>	<i>Zorg voor interne bewustwording van het vastleggen van metadata bij belangrijke veel gebruikte data bronnen.</i>
C: Bewust van data begrijpelijkheid. Wordt veel toegepast ter ondersteuning van de analyse resultaten.		Bij DA verwerken we data en kijken waar mogelijk naar de bijbehorende metadata, desnoods verzamelen we dit specifiek erbij zodat we context van de data beter begrijpen en dit waar nodig mee kunnen nemen in de resultaten.
	<i>Focus:</i>	<i>Verhoog bij de interne data het toepassen van metadata daar waar dit bij data-analyse als gemis ervaren wordt.</i>
D: Maximaal toepassen van data begrijpelijkheid bij elke data-analyse.		Voorafgaand aan DA kijken we altijd naar de begrijpelijkheid van de data en mogelijk aanwezige metadata. Waar nodig wordt dit verder meegenomen in de uitkomsten.
	<i>Focus:</i>	<i>Weet wanneer achtergrond informatie ontbreekt bij data en leg dit vast in het begeleidend schrijven.</i>

## 2.13 Beschikbaarheid - Availability

*De mate waarin gegevens kenmerken hebben waarmee deze kunnen worden opgehaald door geautoriseerde gebruikers en/of toepassingen in een specifieke gebruikscontext. In hoeverre is de data direct gereed of is preparatie nodig om geanalyseerd te kunnen worden.*

A: Er is data welke niet geschikt is voor analyse. We kunnen het niet aanpassen.		Voor het uitvoeren van een data-analyse zijn we niet in staat om verschillende data typen en gegevenskenmerken klaar te zetten voor analyse, kosten kunnen hierbij een reden zijn.
----------------------------------------------------------------------------------	--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



	<i>Focus:</i>	<i>Verhoog kennis van aanwezige data typen binnen de organisatie, weet welke omgezet moeten worden voor efficiënte analyse en hoe deze geschikt gemaakt kunnen worden voor analyse.</i>
B: Data kan, indien nodig, aangepast worden voor analyse, er is geen toegangsbeperking.		Voor het uitvoeren van een data-analyse dienen we verschillende data typen en gegevenskenmerken te prepareren zodat daarna analyse mogelijk is. Analist heeft daarmee toegang tot de data.
	<i>Focus:</i>	Automatiseer het omzetten van data naar een geschikt formaat, daarmee wordt het altijd beschikbaar voor analyse.
C: Eenvoudige data preparatie voor data-analyse		Voor analyse is de data altijd beschikbaar en direct klaar om toegepast te worden bij analyse.
	<i>Focus:</i>	Leg vast hoe de beschikbaarheid bij verschillende data wordt verkregen zodat dit herhaald kan worden.

#### 2.14 Transporteerbaar - Portability

*De mate waarin gegevens kenmerken hebben waarmee deze kunnen worden geïnstalleerd, vervangen of verplaatst van het ene systeem naar het andere, met behoud van de bestaande kwaliteit in een specifieke gebruikcontext.*

A: data niet transporteerbaar		De data nodig bij het uitvoeren van data-analyse is niet te verplaatsen.
	<i>Focus:</i>	<i>Bij data omvang issues, verhoog capaciteit. Verhoog het inzicht waarom het verplaatsing moeilijk is.</i>
B: data verplaatsing vraagt om kennis.		De data nodig bij het uitvoeren van data-analyse is nauwelijks / niet makkelijk te verplaatsen. tussenkomst van specialist is noodzakelijk (databases)
	<i>Focus:</i>	<i>Bij technische beperking achterhaal wat nodig (kennis) is om dit wel te kunnen. Laat indien mogelijk consultant een zodanige aanpassing doen dat transporteerbaarheid vereenvoudigd wordt.</i>
C: data eenvoudig verplaatsbaar		De data nodig bij het uitvoeren van data-analyse is eenvoudig te verplaatsen.
	<i>Focus:</i>	<i>Leg vast hoe moeilijk verplaatsbare data toch verplaatst kan worden – instructie van consultant.</i>

#### 2.15 Herstelbaarheid – Recoverability (Raw data)

*De mate waarin gegevens kenmerken hebben die het mogelijk maken om een bepaald niveau van bewerkingen en kwaliteit te behouden, zelfs in het geval van een storing, in een specifieke gebruikcontext (kan ruwe data worden hersteld waarop analyse heeft plaats gevonden).*

A: Herstelbaarheid data is onbelangrijk.		We kijken niet naar mogelijke herstelbaarheid van data welke is gebruikt bij data-analyse, analyse resultaat is belangrijk.
	<i>Focus:</i>	Snapshot methodes zijn bij uitstek geschikt voor het terug halen van data, bepaal een acceptabele snapshot frequentie. Informeer bij database naar
B: Herstelbaarheid is belangrijk, we vertrouwen op de back-up kwaliteit.		We houden soms rekening met de mogelijkheid tot herstel van ruwe data welke is gebruikt bij data-analyse. We gebruiken waar nodig en indien mogelijk kopieën van data. Bij databases gaan we uit van een goed backup proces te hebben zodat we na een eventuele restore waarschijnlijk dezelfde data hebben voor analyse.

	<i>Focus:</i>	<i>Beschrijf een policy voor het gebruik van data en de herstelbaarheid zodat reproduceren van de analyse mogelijk is. Leg eventuele gebruikte query's vast.</i>
C: Herstelbaarheid is belangrijk, dit in een verwerkingsproces vastgelegd.		We gebruiken altijd kopieën van bestanden bij data-analyse en bewaren deze kopieën altijd (een vaste periode) ongeacht welke analyse zodat bij herhaling de data beschikbaar is. Bij realtime data-analyse weten we dat we een goed backup proces hebben, we zijn er zeker van dat we na een eventuele restore exact dezelfde data hebben voor analyse.
	<i>Focus:</i>	<i>Leg bij nieuwe analyses vast hoe om te gaan met herstelbaarheid.</i>

## 2.16 Reproduceerbaar - Reproducible (Analyse proces)

*De mate waarin gegevens kenmerken hebben die het mogelijk maken om een bepaald niveau van bewerkingen en kwaliteit te behouden, zelfs in het geval van een storing, in een specifieke gebruikscontext (vastleggen van bijvoorbeeld gebruikte query, metadata beschrijving, eventuele gemaakte specifieke instellingen binnen een tool).*

A: Geen vastgelegde instructies en of kennis.		We kijken niet naar mogelijke reproduceerbaarheid van de analyse, elke analyse is zoeken naar een analyse methode, deze wordt niet bewaard.
	<i>Focus:</i>	<i>Verhoog analyse methode kennis bij de data analist</i>
B: Geen vastgelegde instructies, individuele herinnering en kennis is reproduceerbaarheid mogelijk.		We kijken niet naar mogelijke reproduceerbaarheid van de analyse, bij onze herhalende analyses weet de data analist de toegepaste analyse methode, dit is niet specifiek vastgelegd.
	<i>Focus:</i>	<i>Bij belangrijke en of herhalende analyses vast leggen welke data is gebruikt en verwerkt in de analyse.</i>
C: Deels vastgelegde instructies.		Afhankelijk van de uit te voeren data-analyse is de methode van analyse vastgelegd zodat exacte herhaling van de analyse mogelijk is.
	<i>Focus:</i>	<i>Stel regels op wat er bij analyse vastgelegd moet worden zodat reproduceerbaarheid geborgen wordt.</i>
D: Protocol aanwezig.		We bewaren altijd de analyse methode en bijbehorende informatie zodat exacte herhaling van de analyse mogelijk is.
	<i>Focus:</i>	<i>Aanwezige audits bewaken reproduceerbaarheid van processen.</i>

## Bijlage 14, Transcriptie documenten van drie interviews

De drie transcripties zijn om privacy reden alleen voor studie begeleiding beschikbaar gesteld.

## Bijlage 15, 2e iteratie, interview medewerker MKB

Algemene info	Samengevat de feedback
<b>Organisatie 1</b> Online interview: Productiebedrijf met 43 medewerkers. Interview met functionaris: Supply chain manager Opleidingsgraad MBO 25 jaar werkzaam.	Organisatie kent een eenvoudige toepassing van DA. Het is vooral gericht op proces informatie ophalen en analyseren. Daaruit wordt informatie gehaald voor inkoop en financiën. Het concept is lastig, inleiding te summier en vragen vaak onduidelijk verwoord. De structuur van het model is, na interview, duidelijk. Via stellingen het antwoord in een matrix plaatsen en vervolgens aandachtspunten ter verbetering. Toepasbaarheid binnen de organisatie, nee. De organisatie is te klein en het model te lastig om toe te passen.
<b>Organisatie 2</b> Online interview: Productiebedrijf met 150 medewerkers. Interview met functionaris: Data en applicatie beheerder Opleidingsgraad Universiteit half jaar werkzaam bij deze organisatie. (Daarvoor 13,5 jaar als business consultant.)	Organisatie kent een eenvoudige toepassing van DA. Het is vooral gericht op proces informatie ophalen en analyseren. Informatie wordt ingezet voor productie en logistiek Huidige inleiding is veel te summier, deel van de stellingen te onduidelijk. Daarbij was de formulering niet duidelijk of het onderscheid tussen de stellingen te weinig waardoor keuze maken moeilijk was. De structuur van het gehele model was duidelijk. Over toepasbaarheid binnen het MKB: "dit is tooling voor een consultant". "Voor het MKB om als zelfstandig tool te gebruiken is het niet simpel genoeg".
<b>Organisatie 3</b> Interview op locatie bij: Productie bedrijf met 14 medewerkers. Interview met functionaris: directeur. Opleidingsgraad HBO 22 jaar werkzaam.	Organisatie kent een eenvoudige toepassing van DA. Het is vooral gericht op inkoop behoefte bepalen en uren registratie. Het model werd in eerste instantie als lastig ervaren, ook hier werd de beperkte inleiding als een oorzaak aangegeven. De nodige vragen besproken gericht op onduidelijkheden in de stellingen. Algemeen de toepasbaarheid werd ingezien, echter niet binnen eigen organisatie. Vooral op de aandachtsgebieden zou meer kennis nodig zijn. "Volgens mij zou je hier eerst meer kennis van moeten hebben of kennis inkopen via een consultant ".

## Bijlage 16, Overzicht van feedback en aanpassingen uit tweede iteratie.

Gegroepeerd naar algemene feedback, dimensies, matrix en aandachtsgebieden. De concrete aanpassingen zijn in bijlage 11 en 14 te lezen.

### Algemene feedback

Reactie van Respondent 1, 2 of 3	Uitgevoerde aanpassing
2 Inleiding moet duidelijker. Hoe de matrix invullen verder uitschrijven.	Inleiding in zijn geheel herschreven en opmerkingen verwerkt.
2 Kolom JA/Nee bij de stellingen geeft verwarring.	Kolom verwijderd, was nog uit oude opzet.
2 Resultaat matrix aan het begin i.p.v. tussen de stellingen en aandachtsgebieden in.	Resultaat matrix is verplaatst. Staat nu voor de stellingen.
2 Inleidende alinea bij datakwaliteit gaf verwarring.	Alinea verwijderd en informatie verwerkt in de algehele inleiding.
2 Opmerking op maturity level beschrijvingen.	Derde level aangepast bij gegevensbeschikbaarheid. Beschrijving bij gegevenskwaliteit toegevoegd.
2 Had een opmerking over de focus beschrijvingen bij aandachtsgebieden de opmerking dat deze goed gecontroleerd zouden moeten worden.	Alle focus beschrijvingen zijn opnieuw kritisch doorgenomen.
2 In gehele document consequent toepassen van DA i.p.v. data-analyse.	Gehele document gecontroleerd en aangepast
2 Zowel Nederlands als Engelse term bij gegevenskwaliteit.	Engelse term verwijderd.
3 Inleiding was erg beknopt.	Deze is in zijn geheel herschreven.
3 Onduidelijk was of er een maximale score behaald moest worden.	Informatie hierover in de inleiding verwerkt.
3 Inleiding bij Gegevenskwaliteit geeft onduidelijkheid.	Deze informatie verplaatst naar de algemene inleiding en uitgebreider beschreven.
3 Aandachtsgebieden concreet in acties omzetten wordt als lastig ervaren.	Alle aandachtsgebied omschrijvingen en focus beschrijvingen zijn opnieuw kritisch doorgenomen. Punt 1.2 geen vier maar drie aandachtsgebieden. 1.9 focus C anders geformuleerd. 2 beschrijving iets concreter geformuleerd. 14 focus A anders geformuleerd. 2.16 focus D anders geformuleerd. Verder valt er niet veel op aan te passen om het model breed toepasbaar te houden.

## Dimensies

	Reactie van Respondent 1, 2 of 3	Uitgevoerde aanpassing
1.1 Beperking in interne gegevensgebruik	1 Lastig welke stelling het best past. Gebruikte woorden geven verwarring.	Zowel omschrijving als de stellingen aangepast.
1.2 Beperking in externe gegevensgebruik	1 Nu geen gebruik van externe data, mogelijk gaan we dat doen. Formulering nu is lastig om keuze te maken. 2 Wij gebruiken geen externe bronnen, stelling A. Maar dat zegt niks over stelling D, de toegang. 3 Stellingen leken te veel op elkaar.	Stellingen anders geformuleerd
1.3 Niet bestaande of ontbrekende gegevens	2 niet voldoende expliciet beschreven dat het om interne data gaat 3 de omschrijving riep vragen op.	Omschrijving aangepast. Daarnaast bij de stellingen 'data-analyse' afgekort als DA.
1.4 <u>Timing en synchronisatie van gegevens</u>	1 de formulering bij stelling D onduidelijk 2 Stelling A en B complementair.	Alle stelling herzien qua formulering
1.5 <u>Gegevensfrequentie</u>	1 tweede stelling korte vraag, enige onduidelijkheid.	Onduidelijkheid beantwoord. Stelling 1, 2, en 4 anders geformuleerd.
1.6 Gegevensformat, standaarden en specificaties	1 gezien de situatie kon er geen stelling gekozen worden. 2 Formulering dusdanig dat er geen keuze gemaakt kon worden. 3 Stellingen lastig te begrijpen.	Stelling 1, 3, en 4 anders geformuleerd.
1.8 Toegang tot gegevens, opslag en verwerking	1 Stellingen meer tool gericht dan gericht op directe-indirecte toegang tot data.	Stelling 1, 2, 3, en 4 anders geformuleerd.
2.1 Nauwkeurigheid	1 twijfel over de stellingen, geen concrete opmerking	Stellingen niet gewijzigd
2.6 Toegankelijkheid	1 algemene opmerking over wel / geen data toegang	Stellingen niet gewijzigd
2.7 Nakomen	1 woord 'Exact' lastig 2 twee stellingen welke gekozen kon worden.	Stelling 1, 2, 3, en 4 anders geformuleerd.
2.8 Vertrouwelijkheid	2 drie stellingen welke gekozen kon worden.	Stelling 1, 2, 3, en 4 anders geformuleerd. O.a. 'collega's' gewijzigd in 'mensen'
2.11 Traceerbaarheid	2 omschrijving niet geheel duidelijk	Omschrijving aangepast.

2.15 Herstelbaarheid	2 omschrijving niet geheel duidelijk	Omschrijving aangepast.
2.16 Reproduceerbaar	2 omschrijving niet geheel duidelijk	Omschrijving aangepast.

### Matrix

Reactie van Respondent 1, 2 of 3	Uitgevoerde aanpassing:
1 Over Maturity level 0 en de maturity levels in het algemeen "Wat ik me alleen afvraag is of het de bedoeling is om op elk punt de maximale score te krijgen en wat doet die level 0 kolom daar?"	In de inleiding verwoordt.
1 Lege levels (kolommen) bij gegevenskwaliteit, onduidelijke de functie. Dit lijkt geen functie te hebben.	Alle capabilities achter elkaar gezet met maturity level van 0 t/m 5.
1 Aandachtsgebieden. 1.1 geen actie beschreven	Focus omschrijving aangepast.
2 Indeling van Merino naar Inherent, Inherent en Systeem en Systeem in de matrix gaven verwarring.	Deze zijn verwijderd.

### Aandachtgebieden

Aandachtgebieden	Reactie van Respondent 1, 2 of 3	uitgevoerde aanpassing
1.1 beperking in interne gegevensgebruik	2 Beschrijving bij A is meer een vaststelling, daagt niet uit tot actie.	Anders geformuleerd.
1.9 Voorbereidingstijd van gegevens uit andere bronnen	1 geen actie beschreven bij C	Bij C "Er is.." vervangen door "Zorg voor.."
2.10 Precisie	1 opmerking dat verdieping moeilijk is voor hen	Geen aanpassing gedaan. Dit is gerelateerd aan omvang MKB.
2.11 Traceerbaarheid	1 Beschrijving bij C is meer een opmerking, geen actie	Bij C "We hebben vastgelegd wanneer we.." vervangen door "Leg vast wanneer.."

## Bijlage 17, Het FAM model na twee iteraties.

(stap 4 van het DSR-model, Peffers)

### Inleiding

Voor u ligt een zogenaamd maturity model. Een model wat gebruikt wordt om de volwassenheid van processen binnen organisaties vast te leggen op een vaste en meetbare wijze. In welke mate heb je op een schaal van totaal niet tot aan volledige implementatie een proces binnen een organisatie.

### Maturity modellen en het MKB

Dit onderzoek is gefocust op het MKB omdat daar nog weinig kennis bekend is over het toepassen van maturity modellen. Volgens diverse literatuur bronnen wordt het bij het MKB nog weinig toegepast omdat deze bedrijven vaak kenmerkend klein zijn en vervolgens niet de kennis of tijd bezitten om het uit te voeren hebben.

Er zijn verschillende type maturity modellen. Veel type maturity modellen bevatten een vast niveau welke je helemaal moet doorlopen wil je een niveau hoger kunnen behalen en meer volwassenheid willen binnen een proces. Los van de vraag of je dat helemaal wilt behalen is het behalen van een hogere volwassenheid bij deze 'vast niveau' modellen vrij lastig. In dit onderzoek is daarom gekozen voor het Focus Area Model (FAM). Kenmerkend zijn de kleinere stappen die je kunt nemen en daarmee eenvoudiger te behalen. Dit maakt het wellicht juist geschikt voor het MKB. Het FAM model in dit onderzoek is specifiek ontwikkeld en gericht op het Data Analytics (DA) proces. Dit onderzoek wil meten in hoeverre dit model voor het MKB een goede methode is om de mate van volwassenheid van processen te meten en te verhogen.

Eenmaal dit FAM model toegepast wordt het visueel zichtbaar op welke punten het proces zou kunnen worden verbeterd. Als snel zal dan de vraag komen: 'hoe dan en waar dan?'. Daarom is dit onderzoek aangevuld met aandachtsgebieden welke bij deze vragen ondersteuning biedt.

Dit model richt zich op een tweetal belangrijke gebieden van DA namelijk gegevensbeschikbaarheid en datakwaliteit. Na de resultaat matrix begint het met een aantal stellingen. Het duurt ongeveer 30 minuten om deze te doorlopen.

In de volgende paragrafen staan achtereenvolgens een toelichting op de toepassing van de stellingen en de aandachtsgebieden uitgewerkt die doorlopen kunnen worden.

### De stellingen:

Zowel gegevensbeschikbaarheid als datakwaliteit bestaan ieder uit een aantal kenmerken, dimensies genaamd. Voor elke dimensie is er een aantal stellingen geformuleerd en per dimensie dient een stelling gekozen te worden welke het meest van toepassing is binnen de organisatie.

Zo is er bijvoorbeeld bij gegevensbeschikbaarheid een dimensie (kenmerk) "1.1 beperking in interne gegevensgebruik":



1.1 beperking in interne gegevensgebruik = de dimensie	
In welke mate heeft data analist toegang tot benodigde interne data. = korte uitleg van de dimensie	
We kennen een te beperkte toegang tot data welke nodig is voor data analyse. Analyse is niet mogelijk. = stelling 1	A
We hebben beperkte toegang tot data, op basis wat beschikbaar is voeren we data analyse uit. = stelling 2	B
Verschillend per dataset hebben we volledige of beperkte toegang. = stelling 3	C
Als data analist hebben we volledige toegang tot alle interne data. = stelling 4	D

= mogelijk  
antwoord  
waardes

Je ziet boven aan de dimensie “beperking in interne gegevensgebruik” met daaronder (in het zelfde vak) een korte uitleg van de dimensie. Aansluitend daaronder in dit voorbeeld een viertal stellingen (stelling 1 t/m 4) en per stelling een mogelijke stelling in de vorm van een letter, A, B, C of D. Het aantal letters is verschillend per dimensie, soms zijn er maar drie of juist vijf stellingen.

Het is de bedoeling dat je per dimensie een best passende stelling kiest en vervolgens de waarde daarachter overneemt in de resultaat matrix. Deze matrix is direct na deze inleiding geplaatst en ziet er als volgt uit voor gegevensbeschikbaarheid (deels weergegeven). De waardes A, B, C en D zijn reeds geplaatst onder de maturity levels (0 t/m 12). Level 0 vertegenwoordigd het in zijn geheel niet aanwezig zijn van dat kenmerk. Level 12 vertegenwoordigd het maximaal en optimaal gebruik van dat kenmerk voor DA. De onderlinge verschillen per dimensie en onder welke maturity level staat welke letter is op basis van ervaring en argumentatie ingedeeld. Hoe belangrijker het kenmerk hoe eerder deze beginnen met A in de maturity level kolom 1.

		Maturity Level	0	1	2	3	4	5	6	7	8	9	10	11	12
		Focus Area (maturity <u>doelen</u> )													
<b>1</b>	<b>Gegevensbeschikbaarheid</b>														
1.1	beperking in interne gegevensgebruik			A	B					C			D		
1.2	beperking in externe gegevensgebruik				A	B				C				D	
1.3	niet bestaande of ontbrekende gegevens						A				B			C	
1.4	timing en synchronisatie van gegevens				A			B				C			D

Invullen van de matrix:

Als je bij dimensie 1.1 hebt gekozen voor de stelling welke het best de situatie beschrijft, bijvoorbeeld: “We hebben beperkte toegang tot data, op basis wat beschikbaar is voeren we data-analyse uit”, dan staat daarachter de ‘B’.

Vervolgens is het de bedoeling om in het resultaat matrix vanaf maturity level kolom ‘0’ tot en met de letter welke is gekozen in te kleuren. Dit geldt voor elke dimensie.

Ingevuld kan het er als volgt uit komen te zien:

		Maturity Level	0	1	2	3	4	5	6	7	8	9	10	11	12
		Focus Area (maturity <u>doelen</u> )													
<b>1</b>	<b>Gegevensbeschikbaarheid</b>														
1.1	beperking in interne gegevensgebruik			A	B					C			D		
1.2	beperking in externe gegevensgebruik				A	B				C				D	
1.3	niet bestaande of ontbrekende gegevens						A				B			C	
1.4	timing en synchronisatie van gegevens				A			B				C			D

Eenmaal alles ingekleurd kan je eenvoudig zien op welke onderdelen ( de dimensies) je proces meer en of minder een mate van volwassenheid bezit. Hoe meer volwassenheid hoe beter de kwaliteit van

de toegepaste data-analyses. Op alle dimensies een maximale maturity level is niet realistisch. Het is een meet moment wat meer en minder ontwikkelde onderdelen zichtbaar maakt en waarop je vervolgens keuzes kunt maken.

### De aandachtsgebieden:

Per dimensie is er vervolgens ook een aandachtsgebied beschreven. Zodra je de resultaat matrix hebt ingevuld en hebt besloten om bij één of een aantal dimensies een hogere volwassenheid te willen behalen, kan je gebruik maken van de beschreven aandachtsgebieden. Deze ondersteunen je vervolgens met het geven van gerichte acties welke uitgevoerd kunnen worden of waaraan gedacht moet worden bij uitvoering.

De aandachtsgebieden van bijvoorbeeld “1.1 beperking in interne gegevensgebruik” ziet er als volgt uit:

1.1 beperking in interne gegevensgebruik	
<i>De mate waarin de data-analist tijdig en in voldoende mate over interne data kan beschikken.</i>	
<b>A</b>	<i>Focus: Zorg dat je meer inzicht krijgt in het belang van (hoeveelheid) data in relatie tot kwaliteit van DA. De toegang tot interne data voor DA functie/taak meer openstellen en of meer vertrouwen krijgen in de data analist.</i>
<b>B</b>	<i>Focus: Interne data toegang verbeteren door zoveel mogelijk interne data toegankelijk maken voor data analist. Bestanden in persoonlijke mappen op algemeen toegankelijke locaties plaatsen. Meer vertrouwen in de data analist krijgen en nemen. Zorg dat de analist overzicht krijgt van aanwezige interne data bronnen.</i>
<b>C</b>	<i>Focus: Volledige medewerking en aanlevering van data aan de data analist bij verzoeken. Eventueel rechten uitbreiden op databases, zorg dat data analist kennis heeft van aanwezige tabellen en minimaal toegang heeft.</i>
<b>D</b>	<i>Focus: Blijft organisatie breed bewust van data hoeveelheid in relatie tot kwaliteit van data analyse.</i>

Kies in dit overzicht dezelfde letter zoals deze bij de stellingen is gekozen. Dus bij 1.1 is in het voorbeeld hierboven gekozen voor de waarde ‘B’. De omschrijving achter ‘B’ bevat concrete acties uit te voeren en zo te helpen om een betere volwassenheid te creëren bij data-analyse en richting een maturity ‘C’ te gaan.

In onderstaand model is een aantal begrippen gebruikt. Om verwarring te voorkomen worden eerst deze begrippen kort beschreven:

**DA:** afkorting van data-analyse.

**Ruwe data:** data zoals deze is direct na verzameling en geen verdere verwerking heeft ondergaan.

**Dataset:** een hoeveelheid ruwe data waarop een DA wordt uitgevoerd.

**Interne data:** data verzameld en opgeslagen welke afkomstig is vanuit de eigen organisatie en toegepast bij DA.

**Externe data:** data afkomstig van online databases, klanten, leveranciers, etc. en toegepast bij DA.

**Focusgebied:** in dit onderzoek zijn dat gegevensbeschikbaarheid en datakwaliteit.

**Dimensies:** de kenmerken per focusgebied, elke dimensie bevat in dit onderzoek een aantal stellingen.

**Metadata:** een begeleid schrijven behorend bij een dataset. Het geeft meer informatie over de context van de ruwe data weer zodat deze beter geïnterpreteerd en verwerkt kan worden.

**Tijdsynchronisatie:** het synchroon lopen van verschillende klokken, vooral bij het vastleggen van gebeurtenissen (data logging). Dit maakt het mogelijk om van verschillende data bronnen de tijdstipmomenten te koppelen

## Resultaat Matrix

		Maturity Level	0	1	2	3	4	5	6	7	8	9	10	11	12
		Focus Area (maturity doelen)													
1	<b>Gegevensbeschikbaarheid</b>														
1.1		beperking in interne gegevensgebruik		A	B					C			D		
1.2		beperking in externe gegevensgebruik			A	B							C		
1.3		niet bestaande of ontbrekende gegevens					A				B			C	
1.4		timing en synchronisatie van gegevens			A			B				C			D
1.5		gegevensfrequentie			A		B		C				D		
1.6		Gegevensformat, standaarden en specificaties				A		B			C			D	
1.7		Onzekerheid, betrouwbaarheid en nauwkeurigheid				A		B				C			D
1.8		Toegang tot gegevens, opslag en verwerking					A	B		C			D		
1.9		Vorbereidingstijd van gegevens uit andere bronnen							A		B			C	
				kwaliteit DA niet of nauwelijks mogelijk			kwaliteit DA beperkt mogelijk			kwaliteit DA geoptima- liseerd			Kwaliteit DA volledig mogelijk		

2	Maturity Level					0	1	2	3	4	5
	Gegevenskwaliteit										
2.1	Nauwkeurigheid					A	B	C			
2.2	Volledigheid					A	B	C			
2.3	Consistentie					A	B	C	D		
2.4	Geloofwaardigheid					A	B	C	D		
2.5	Actueel					A	B	C	D		
2.6	Toegankelijkheid					A	B	C	D	E	
2.7	Nakomen					A	B	C	D		
2.8	Vertrouwelijkheid					A	B	C			
2.9	Efficiëntie					A	B	C	D		
2.10	Precisie					A	B	C			
2.11	Traceerbaarheid					A	B	C			
2.12	Begrijpelijkheid					A	B	C			
2.13	Beschikbaarheid					A	B	C			
2.14	Transporteerbaarheid					A	B	C			
2.15	Herstelbaarheid					A	B	C			
2.16	Reproduceerbaarheid					A	B	C	D		
						Bepaalde gegevens - kwaliteit	Betere gegevens - kwaliteit				

# Maturity Model

## Gegevensbeschikbaarheid

Gegevensbeschikbaarheid is gedefinieerd als 'de graad van gemak voor gebruikers om gegevens en gerelateerde informatie te verkrijgen'. Vanuit een meer technische invalshoek kan je het ook zien als 'de mate waarin data continue beschikbaar is' maar dat richt zich meer op technische oplossingen en valt buiten dit onderzoek.

In het algemeen zijn belangrijke vragen 'welke gegevens verzamelen we (variety)', 'hoe verzamelen we de steeds maar in snelheid toenemende data (velocity)' en 'hoeveel verzamelen we (volume)' waarbij de antwoorden sterk de gegevensbeschikbaarheid bepalen. Hieronder staan negen dimensies (punt 1.1 t/m 1.9) van gegevensbeschikbaarheid zoals deze in de literatuur zijn verwoord en deels aangepast om geschikt te zijn in dit onderzoek. Per dimensie staat een aantal stellingen waarbij een stelling gekozen moet worden welke het best de situatie bij de organisatie beschrijft.

1.1 beperking in interne gegevensgebruik		
In welke mate heeft data analist toegang tot interne data zodat diverse analyses uitgevoerd kunnen worden.		
We kennen een te beperkte toegang tot interne data ook voor de analist welke mogelijk data nodig heeft voor DA. Kwantitatief te weinig data voor DA.		<b>A</b>
We kennen een te beperkte toegang tot interne data ook voor de analist welke mogelijk data nodig heeft voor DA. De hoeveelheid data toegang is ondanks de beperkte toegang dermate veel dat op basis van wat beschikbaar is DA wordt toegepast.		<b>B</b>
De data analist heeft meer toegang tot data gekregen. De toegang is verschillend per dataset, volledige of beperkt. Diverse analyses zijn mogelijk.		<b>C</b>
Als data analist hebben we volledige toegang tot alle interne data. Alle analyses zijn in mogelijk.		<b>D</b>

1.2 externe gegevensgebruik gebruik		
In welke mate beschikt de data analist over externe data of kan deze er over beschikken.		
We gebruiken geen externe data bronnen		<b>A</b>
We gebruiken externe data bronnen minimaal bij DA.		<b>B</b>
We gebruiken externe data bronnen veelvuldig bij DA.		<b>C</b>

1.3 niet bestaande of ontbrekende gegevens		
In hoeverre wordt er voor specifieke DA nog extra data verzameld binnen de organisatie (interne data), naast de normale proces data, om een meer gefundeerdere analyse te kunnen uitvoeren.		
We verzamelen nooit extra data binnen de organisatie specifiek voor toepassing bij DA.		<b>A</b>

We verzamelen incidenteel extra data binnen de organisatie specifiek voor toepassing bij DA.	<b>B</b>
We verzamelen continue extra data binnen de organisatie specifiek voor toepassing bij DA.	<b>C</b>

<b>1.4 timing en synchronisatie van gegevens</b>  <i>Bij data-analyse kan het moment (tijd) waarbij een waarde is vastgelegd belangrijk zijn. Vooral bij het gebruik van meerdere data bronnen is het van belang dat er zekerheid is zodat geregistreerde tijden onderling vergelijkbaar zijn bijvoorbeeld chronologische volgorde van vastgelegde data.</i>	
Bij onze DA speelt tijd en daarmee tijdsynchronisatie geen rol.	<b>A</b>
Bij onze DA speelt tijd een rol. Echter tijdsynchronisatie is niet van belang omdat we geen koppeling hebben met ander data bronnen.	<b>B</b>
Bij onze DA speelt tijd een rol. Van alle data bronnen is bekend dat de tijd synchroon is.	<b>C</b>
Bij onze DA speelt tijd een rol. Van alle data bronnen is bekend dat de tijd synchroon is. We controleren hier expliciet op bij aanvang van DA.	<b>D</b>

<b>1.5 gegevensfrequentie</b>  <i>De frequentie waarop interne data wordt verzameld / vastgelegd. (bijvoorbeeld bij sensoren op productiemachines). Meer beschikbare data kan leiden tot hogere nauwkeurigheid en betere inzichten, meer opslag betekent o.a. meer opslagruimte vereist en hogere kosten.</i>	
We kennen geen frequente datavastlegging, incidenteel verzamelen we data voor DA.	<b>A</b>
We kennen een frequente datavastlegging.	<b>B</b>
We kunnen de frequentie van datavastlegging verhogen zodat we meer data verzamelen (bewust van de mogelijkheid).	<b>C</b>
Een of meerdere data registratie (verzamel) punten leveren met opzet een grotere hoeveelheid data, hier kunnen we gebruik van maken voor DA.	<b>D</b>

<b>1.6 Gegevensformat, standaarden en specificaties</b>  <i>Organisaties kennen ongelijksoortige gegevens en vaak opgeslagen in bestanden geformatteerd volgens verschillende normen. Het kunnen normaliseren / standaardiseren van vast te leggen data voor analyse kan van belang zijn.</i>	
Onze data bestaat met name uit handmatig ingevulde formulieren, voor analyse wordt deze data eerst digitaal gemaakt.	<b>A</b>
Data converteren doen we niet. We kunnen het verwerken zoals het nu is.	<b>B</b>
Als we converteren dan is dat niet op basis van een vast protocol (herhaling van de handeling is moeilijker).	<b>C</b>
We converteren ruwe data en volgens vast protocol zodat exacte herhaling mogelijk is.	<b>D</b>

### 1.7 Onzekerheid, betrouwbaarheid en nauwkeurigheid

*Bedrijfsgegevens kunnen kenmerken hebben welke van invloed zijn op Data Analytics zoals onzekerheid, betrouwbaarheid en nauwkeurigheid. Metadata kan hier bij helpen. Het feit dat analyses in toenemende mate bepalend zijn voor besluitvorming, is het van essentieel belang dat deze data betrouwbaar is.*

Bij DA controleren we onze data niet specifiek op onzekerheid, betrouwbaarheid en of nauwkeurigheid.	<b>A</b>
Incidenteel doen we aan verbetering van data onzekerheid, betrouwbaarheid of nauwkeurigheid.	<b>B</b>
We gebruiken incidenteel metadata gegevens ter verbetering van data onzekerheid, betrouwbaarheid of nauwkeurigheid.	<b>C</b>
We gebruiken veelvuldig metadata voor verbetering van data onzekerheid, betrouwbaarheid en nauwkeurigheid	<b>D</b>

### 1.8 Toegang tot gegevens, opslag en verwerking

*Bij toegang is de vraag of er direct op de productie data gewerkt wordt of dat er voor DA met een kopie gewerkt wordt. Bij dit laatste kan data onbewust aangepast zijn zodat het als kopie opgeslagen kan worden, denk een grote dataset waar alleen maar een klein deel nodig is. Daarnaast kan data in verschillende type databases opgeslagen zijn. Elke heeft zijn voor- nadelen bij verwerking in DA.*

We gebruiken alleen kopieën van de ruwe data bronnen bij data-analyse. Om technische reden gebruiken we kopieën.	<b>A</b>
We gebruiken alleen kopieën van de ruwe data bronnen bij data-analyse. Om veiligheidsreden gebruiken we kopieën.	<b>B</b>
Bij het gebruik van data voor analyse hebben we een mix van zowel data kopieën als ook directe toegang tot ruwe data bronnen.	<b>C</b>
Bij het gebruik van data voor analyse hebben we meestal directe toegang tot de ruwe data bronnen.	<b>D</b>

### 1.9 Voorbereidingstijd van gegevens uit andere bronnen

*Het verkrijgen van data uit andere (intern / extern) niet standaard bronnen vereist voorbereiding voor verwerking. Voor verwerking en interpretatie dient de ontvangen data geanalyseerd te worden voorafgaand aan verdere verwerking binnen DA en eventuele metadata gegevens doorgenomen te worden.*

We gebruiken geen data uit andere data bronnen en daarmee geen voorbereiding van data nodig.	<b>A</b>
We gebruiken data uit andere data bronnen. We voeren voorafgaand aan DA soms een voorbereiding daarop uit.	<b>B</b>
We gebruiken data uit andere data bronnen. We voeren voorafgaand aan DA consequent een controle daarop uit. Dit ter controle zodat er bij analyse geen verassingen meer zijn.	<b>C</b>

## Datakwaliteit

Dit onderzoek hanteert als definitie voor datakwaliteit “de mate waarin de kwaliteit voldoet aan, of deze de kwaliteitseisen overtreft, die gebruikers ervan verwachten”.

De literatuur beschrijft algemene kenmerken, trends over gegevensbeschikbaarheid en gegevenskwaliteit. Een goede data-analyse is sterk afhankelijk van de kwantiteit en kwaliteit van gegevens welke geen van beide gemakkelijk beschikbaar zijn binnen het MKB in vergelijking tot grote ondernemingen, het aantal interne gegevensbronnen is veelal beperkt.

Hieronder zestien dimensies (punt 2.1 t/m 2.16) van datakwaliteit zoals deze in de literatuur zijn verwoord en deels aangepast om geschikt te zijn in dit onderzoek. Per dimensie een aantal stellingen waarbij een stelling gekozen moet worden welke het best de situatie bij de organisatie beschrijft.

<b>2.1 Nauwkeurigheid</b>  <i>De nauwkeurigheid geeft aan hoeveel de meetwaarde of informatie van de werkelijkheid af zit. Nauwkeurigheid is te bepalen door de meting te vergelijken met een meting waarvan de nauwkeurigheid bekend is.</i>		
Incidenteel komt het voor dat voorafgaand aan DA de data op nauwkeurigheid wordt beoordeeld.		<b>A</b>
Structureel wordt er voorafgaand aan DA de data op nauwkeurigheid beoordeeld.		<b>B</b>
We kennen data audits waarbij specifiek gecontroleerd word op nauwkeurigheid		<b>C</b>

<b>2.2 Volledigheid</b>  <i>Bepalen welke datavelden je nodig hebt voor de analyse. Daarnaast per dataveld controle of het daadwerkelijk gegevens bevat.</i>		
We controleren voorafgaand aan DA nooit op de volledigheid van datasets.		<b>A</b>
Afhankelijk van de DA controleren we de volledigheid van datasets (incidentele controle).		<b>B</b>
Bij elke DA controleren we de volledigheid van datasets (structurele controle).		<b>C</b>

<b>2.3 Consistentie</b>  <i>De gegevens mogen geen hiaten bevatten in het tijdspad van analyse (tijdconsistentie), geen data duplicaten en vrij van tegenstrijdigheden. Hebben meet sensoren bijvoorbeeld continue goed gefunctioneerd en data constant aangeleverd, zijn handmatige formulieren door iedereen op een zelfde wijze ingevuld.</i>		
Het is mogelijk dat datasets afwijkende consistentie hebben, we controleren dit in zijn geheel niet.		<b>A</b>
Het is mogelijk dat datasets afwijkende consistentie hebben, afhankelijk per analyse controleren we dit.		<b>B</b>
Het is mogelijk dat datasets afwijkende consistentie hebben, bij elke analyse controleren we dit handmatig, niet volgens een protocol.		<b>C</b>

Het is mogelijk dat datasets afwijkende consistentie hebben, bij elke analyse controleren we dit geautomatiseerd.	<b>D</b>
-------------------------------------------------------------------------------------------------------------------	----------

#### 2.4 Geloofwaardigheid

*De mate waarin gegevens kenmerken hebben die in een specifieke gebruikscontext als waar en geloofwaardig (authentiek) kunnen worden beschouwd en de verzameling op een juiste wijze heeft plaatsgevonden.*

We controleren nooit de authenticiteit (databron) van de ruwe data.	<b>A</b>
We controleren bij twijfels de authenticiteit (databron) van de ruwe data.	<b>B</b>
Waar mogelijk zullen we ruwe data controleren op authenticiteit.	<b>C</b>
Waar mogelijk zullen we ruwe data controleren op authenticiteit en controleren waar nodig hoe de data verzameling heeft plaats gevonden.	<b>D</b>

#### 2.5 Actueel

*Bevat de data kenmerken van de juiste leeftijd binnen een specifieke analyse context.*

We gaan er vanuit dat data actueel is bij aanvang van een analyse, er vindt geen specifieke controle plaats.	<b>A</b>
Zodra er vragen zijn welke na analyse ontstaan over actualiteit van de data worden deze besproken.	<b>B</b>
We doen steekproeven bij data waarbij gecontroleerd wordt of deze actueel is bij aanvang van een analyse.	<b>C</b>
We voeren altijd vaste handelingen uit bij data waarbij gecontroleerd wordt of deze actueel is bij aanvang van een analyse.	<b>D</b>

#### 2.6 Toegankelijkheid

*De mate waarin gegevens toegankelijk zijn in een specifieke gebruikscontext, met name door mensen die ondersteunende technologie of speciale configuratie nodig hebben vanwege een toegangsbeperking.*

Als data analist heb ik standaard geen of beperkte toegang tot data.	<b>A</b>
Als data analist krijg ik vooraf specifieke toegangsrechten.	<b>B</b>
Als data analist heb ik via een kopie toegang tot data, real-time analyses niet mogelijk.	<b>C</b>
Als data analist heb ik via een kopie toegang tot data, deze data is gesynchroniseerd, real-time analyses zijn mogelijk.	<b>D</b>
Als data analist heb ik toegang tot alle centrale data.	<b>E</b>

#### 2.7 Naleven



<i>De mate waarin gegevens kenmerken hebben die voldoen aan geldende normen, conventies of voorschriften en soortgelijke regels met betrekking tot gegevenskwaliteit in een specifieke gebruikscontext.</i>		
Er zijn voorschriften (wettelijke of vanuit de branche) m.b.t. onze data en de verwerking. We voldoen hier niet aan.		<b>A</b>
Er zijn voorschriften (wettelijke of vanuit de branche) m.b.t. onze data en de verwerking. We voldoen hier aan.		<b>B</b>
Naast externe voorschriften hebben we informele interne regels hoe we met data en dataverwerking omgaan.		<b>C</b>
Naast externe voorschriften hebben we formele interne vastgelegde regels hoe we met data en dataverwerking omgaan.		<b>D</b>

2.8 Vertrouwelijkheid		
<i>Hoe betrouwbaar is de data binnen de organisatie. In welke mate is het toegankelijk en is deze interpreteerbaar zonder kennis van de context.</i>		
Data gebruikt voor analyse is toegankelijk voor mensen binnen de organisatie		<b>A</b>
Data gebruikt voor analyse is voor een beperkte aantal mensen toegankelijk. Zonder kennis is deze data interpreteerbaar.		<b>B</b>
Data gebruikt voor analyse is voor een beperkte aantal mensen toegankelijk. Alleen met kennis is deze data interpreteerbaar.		<b>C</b>

2.9 Efficiëntie		
<i>Het kunnen aanpassen van data zodat dit de prestatie van analyse verbetert bijvoorbeeld het verkleinen van de data omvang (deductie, compressie, ontdubbelen)</i>		
We doen niet aan data efficiëntie, geen deductie, compressie of ontdubbeling.		<b>A</b>
Onze data zou met extra bewerking wellicht eenvoudiger te verwerken zijn bij DA.		<b>B</b>
We passen soms data aan zodat de analyse verwerking verbetert.		<b>C</b>
Structureel wordt de nodige data aangepast om de analyse verwerking beter te laten presteren.		<b>D</b>

2.10 Precisie		
<i>In hoeverre bevat de data waarden welke, afhankelijk van de context, op meerdere wijzen interpreteerbaar is.</i>		
We weten niet of data afhankelijk van de analyse verschillende waarden kan opleveren.		<b>A</b>
We kijken niet naar data of deze afhankelijk van de analyse verschillende waarden kan opleveren		<b>B</b>
We weten van onze data dat deze op meerdere wijzen interpreteerbaar is.		<b>C</b>

2.11 Traceerbaarheid		
<i>Kunnen we controleren op mutaties in de data nadat deze data is gebruikt voor analyse. Zijn we in staat om een analyse nogmaals toe te passen op de data en is de uitkomst gelijk</i>		
Bij DA hebben we nog nooit een reeds uitgevoerde analyse nogmaals exact hetzelfde uitgevoerd over dezelfde dataset.		<b>A</b>
We hebben geen controle op eventuele mutaties in de data nadat we deze hebben gebruikt voor analyse.		<b>B</b>
Data welke we voor analyse gebruiken komt uit een database systeem. Binnen dit systeem vindt er mutatie logging plaats. We zijn in staat om dit te controleren.		<b>C</b>

2.12 Begrijpelijkheid		
<i>In hoeverre hebben gegevens kenmerken die het mogelijk maken om door gebruikers te lezen en te interpreteren, en worden uitgedrukt in geschikte talen, symbolen en eenheden in een specifieke gebruikscontext. Sommige informatie beschikt over metadata, een begeleidend schrijven.</i>		
De waardes in onze ruwe data zijn ook te begrijpen voor een leek		<b>A</b>
De waardes in onze ruwe data zijn na enige analyse te begrijpen voor een leek (eventueel na doornemen metadata)		<b>B</b>
De waardes in onze ruwe data zijn onduidelijk voor een leek ook na enige analyse (en eventueel na doornemen metadata)		<b>C</b>

2.13 Beschikbaarheid		
<i>In hoeverre is de data direct gereed of is preparatie nodig om geanalyseerd te kunnen worden.</i>		
Het komt voor dat we data (bestanden) technisch niet aangepast krijgen zodat we deze kunnen analyseren.		<b>A</b>
Alle data welke we tot op heden gebruiken voor analyse kunnen we verwerken, eventueel na eenvoudige aanpassing (bijv. converteren)		<b>B</b>
De data gebruikt voor analyse kunnen we aanpassen zodat deze beter te verwerken is bij de DA.		<b>C</b>

2.14 Transporteerbaarheid		
<i>De mate waarin gegevens kenmerken hebben waarmee deze kunnen worden geïnstalleerd, vervangen of verplaatst van het ene systeem naar het andere, met behoud van de bestaande kwaliteit in een specifieke gebruikscontext.</i>		
Onze data bevat dusdanige kenmerken (omvang, type, security) dat deze niet te verplaatsen is.		<b>A</b>
Indien data voor DA verplaatst moet worden vraagt dit specialistische kennis		<b>B</b>
Onze data, gebruikt bij DA, kan eenvoudig verplaatst worden indien dat nodig is.		<b>C</b>

2.15 Herstelbaarheid		
----------------------	--	--

<i>De mate waarin ruwe data kan worden hersteld (terug halen van backup naar een specifiek tijdstip) naar de situatie waarop analyse heeft plaats gevonden en exact dezelfde analyse weer uitgevoerd kan worden.</i>		
Het kan zijn dat DA niet nogmaals uitgevoerd kan worden op exact dezelfde ruwe data.		<b>A</b>
We maken van ruwe data waar mogelijk kopieën zekerheidshalve en bewaren deze zodat exacte herhaling mogelijk is.		<b>B</b>
Bij het eventueel moeten herhalen van een analyse zal dit op basis van ervaring van de data analist worden uitgevoerd.		<b>C</b>

## 2.16 Reproduceerbaarheid

*De mate waarin gegevens en analyse proces kenmerken hebben die het mogelijk maken om een bepaald niveau van bewerkingen en kwaliteit te behouden, zelfs in het geval van een storing, in een specifieke gebruikscontext (vastleggen van bijvoorbeeld gebruikte query, metadata beschrijving, eventuele gemaakte specifieke instellingen binnen een tool).*

We kijken niet naar mogelijke reproduceerbaarheid van de analyse, elke analyse is zoeken naar een analyse methode, deze wordt niet bewaard.		<b>A</b>
We kijken niet naar mogelijke reproduceerbaarheid van de analyse, bij onze herhalende analyses weet de data analist de toegepaste analyse methode, dit is niet specifiek vastgelegd.		<b>B</b>
Afhankelijk van de uit te voeren data-analyse is de methode van analyse vastgelegd zodat exacte herhaling van de analyse mogelijk is.		<b>C</b>
We bewaren altijd de analyse methode en bijbehorende informatie zodat exacte herhaling van de analyse mogelijk is		<b>D</b>

## Dimensie verbeteringspunten

### Gegevensbeschikbaarheid

1.1 beperking in interne gegevensgebruik	
In welke mate heeft data analist toegang tot benodigde interne data zodat diverse analyses uitgevoerd kunnen worden.	
<b>A</b>	<i>Focus: Zorg dat je meer inzicht krijgt in het belang van (hoeveelheid) data in relatie tot kwaliteit van DA. De toegang tot interne data voor DA functie/taak meer openstellen en of meer vertrouwen krijgen in de data analist.</i>
<b>B</b>	<i>Focus: Interne data toegang verbeteren door zoveel mogelijk interne data toegankelijk maken voor data analist. Bestanden in persoonlijke mappen op algemeen toegankelijke locaties plaatsen. Meer vertrouwen in de data analist krijgen en nemen. Zorg dat de analist overzicht krijgt van aanwezige interne data bronnen.</i>
<b>C</b>	<i>Focus: Volledige medewerking en aanlevering van data aan de data analist bij verzoeken. Eventueel rechten uitbreiden op databases, zorg dat data analist kennis heeft van aanwezige tabellen en minimaal toegang heeft.</i>
<b>D</b>	<i>Focus: Blijft organisatie breed bewust van data hoeveelheid in relatie tot kwaliteit van DA.</i>

1.2 externe gegevensgebruik	
In welke mate beschikt de data analist over externe data of kan deze er over beschikken.	
<b>A</b>	<i>Focus: Vind mogelijke externe data bronnen. Zorg waar mogelijk voor ingangen zodat bij een data behoefte deze ingang beschikbaar is.</i>
<b>B</b>	<i>Focus: Verkrijg inzicht in zoveel mogelijke externe data bronnen. Tijdig opvragen externe data wanneer nodig, regel indien mogelijk een directe remote toegang.</i>
<b>C</b>	<i>Focus: Inventariseer of er via software een directe koppeling mogelijk is met eigen DA pakketten. Ken de vereisten van de externe bronnen om toegang te behouden (abonnementskosten, periodiek vereiste wachtwoord verversing, regelmatig inloggen, etc).</i>

1.3 niet bestaande of ontbrekende gegevens	
In hoeverre wordt er voor specifieke DA nog extra data verzameld binnen de organisatie (interne data), naast de normale proces data, om een meer gefundeerdere analyse te kunnen uitvoeren.	
<b>A</b>	<i>Focus:</i>

	<i>Overweeg het gebruik van extra (data registratie) sensoren. Digitale sensoren of handmatige registratie gericht op plaatsen in het proces waar DA voor ingezet wordt.</i>
<b>B</b>	<i>Focus: Zorg voor regelmaat van sensor informatie. Weet wanneer sensor registratie uit is gebleven bij langdurige vastlegging. Vervang handmatige vastlegging van data naar geautomatiseerde.</i>
<b>C</b>	<i>Focus: Zorg voor data opslag capaciteit, maak via automatische controles zichtbaar dat er registraties ontbreken, bijvoorbeeld via logging van data opslag frequentie eventuele hiaten vinden.</i>

#### 1.4 timing en synchronisatie van gegevens

*Bij data-analyse kan het moment (tijd) waarbij een waarde is vastgelegd belangrijk zijn. Vooral bij het gebruik van meerdere data bronnen is het van belang dat er zekerheid is zodat geregistreeerde tijden onderling vergelijkbaar zijn, bijvoorbeeld chronologische volgorde van vastgelegde data.*

<b>A</b>	<i>Focus: Analyseer de verzamelde data op het mogelijk vastleggen van tijdsregistraties. Verkrijg inzicht in bij welke DA timing en synchronisatie van belang is.</i>
<b>B</b>	<i>Focus: Zorg dat bij vastlegging van data ook de tijden worden vastgelegd en dat belangrijke sensoren onderling geijkt zijn qua tijd/datum.</i>
<b>C</b>	<i>Focus: Zorg dat bij vastlegging van alle data ook de tijden worden vastgelegd waar nodig en deze onderling geijkt zijn zodat ongeacht de analyse er een vergelijking mogelijk is.</i>
<b>D</b>	<i>Focus: Bewaak de timing en synchronisatie van de sensoren, zorg eventueel voor connectie naar een online time server.</i>

#### 1.5 Gegevensfrequentie

*De frequentie waarop interne data wordt verzameld / vastgelegd. (bijvoorbeeld bij sensoren op productiemachines). Meer beschikbare data kan leiden tot hogere nauwkeurigheid en betere inzichten, meer opslag betekent o.a. meer opslagruimte vereist en hogere kosten.*

<b>A</b>	<i>Focus: Weet welke analyse je wilt uitvoeren en welke informatie nodig is en wat er aan data verzameld kan worden. Bepaal vooraf een verzamel periode.</i>
<b>B</b>	<i>Focus: Zorg voor continue vastleggen van data op vaste meet punten in regelmatig te analyseren bedrijfsprocessen. Plaats digitale sensoren indien mogelijk, minder kans op menselijke fouten, hiaten.</i>
<b>C</b>	<i>Focus: Verhoog de frequentie van data verzameling, denk aan opslag capaciteit. Meer data voor analyse levert meer betrouwbare analyses.</i>
<b>D</b>	<i>Focus: Zet de frequentie van data opslag op maximaal. Plaats indien gewenst meer sensoren.</i>

#### 1.6 Gegevensformat, standaarden en specificaties

*Organisaties kennen ongelijksoortige gegevens en vaak opgeslagen in bestanden geformatteerd volgens verschillende normen.*

<i>Het kunnen normaliseren / standaardiseren van vast te leggen data voor analyse kan van belang zijn.</i>	
<b>A</b>	<i>Focus: Handmatige data registraties waar mogelijk digitaal invoeren/registreren.</i>
<b>B</b>	<i>Focus: Beperk de data invoer tot bekende formats en bij keuze beperk het aantal formats. Bijvoorbeeld alleen in een database of in Excel. Zorg bij aangeleverde externe data dat het format toegepast kan worden bij de analyse.</i>
<b>C</b>	<i>Focus: Zorg dat zoveel als mogelijk data voldoet aan een standaard format en geschikt is voor analyse. Standaardiseer de data formats van data welke frequent voor analyses gebruikt wordt.</i>
<b>D</b>	<i>Focus: Weet hoe diverse data formaten verwerkt en vervolgens toegepast kunnen worden bij analyse. Standaardiseer zoveel mogelijk interne data opslag zodat deze altijd geschikt is voor toekomstige analyse. Gebruik analyse tools welke overweg kunnen met verschillende data formats.</i>

### 1.7 Onzekerheid, betrouwbaarheid en nauwkeurigheid

*Bedrijfsgegevens kunnen kenmerken hebben welke van invloed zijn op Data Analytics zoals onzekerheid, betrouwbaarheid en nauwkeurigheid. Metadata kan hier bij helpen. Het feit dat analyses in toenemende mate bepalend zijn voor besluitvorming, is het van essentieel belang dat deze data betrouwbaar is.*

<b>A</b>	<i>Focus: Bij herhalende analyses verschaft inzicht in mogelijke punten binnen de data waar steekproef controle mogelijk is. Stel voorafgaand aan DA vast hoe betrouwbaar de data kan zijn (bron controle, is data actueel/bijgewerkt).</i>
<b>B</b>	<i>Focus: Controleer vooraf of er metadata beschikbaar is bij de dataset en neem dit door.</i>
<b>C</b>	<i>Focus: Creëer bewustwording binnen de organisatie gericht op data betrouwbaarheid en nauwkeurigheid. Creëer bij data vastlegging de mogelijkheid voor mensen / machines om opmerkingen of zogenaamde Tag's te plaatsen (in metadata) waar twijfels bestaan over betrouwbaarheid of nauwkeurigheid.</i>
<b>D</b>	<i>Focus: Bewaak de audits en maak resultaten kenbaar binnen de organisatie.</i>

### 1.8 Toegang tot gegevens, opslag en verwerking

*Bij toegang is de vraag of er direct op de productie data gewerkt wordt of dat er voor DA met een kopie gewerkt wordt. Bij dit laatste kan data onbewust aangepast zijn zodat het als kopie opgeslagen kan worden, denk een grote dataset waar alleen maar een klein deel nodig is. Daarnaast kan data in verschillende type databases opgeslagen zijn. Elke heeft zijn voor- nadelen bij verwerking in DA.*

<b>A</b>	<i>Focus: Weet welke data kopieën mogelijk aangepast zijn, en doe daarbij een controle met de originele data. Zorg waar mogelijk dat toekomstige kopieën structureel geen data wijzigingen bevatten, ken in ieder geval de probleem kopieën. (denk bijv. aan export van berekende velden)</i>
<b>B</b>	<i>Focus: Zorg voor een structureel mechanisme (Checksums) wat de kopie data controleert met het origineel.</i>
<b>C</b>	<i>Focus:</i>

	<i>Vergroot de kennis van databases, vooral het koppelen van verschillende databases en tabellen. Gebruik eventueel software welke hierbij ondersteunt.</i>
<b>D</b>	<i>Focus: Zorg bij kopie data, waarop belangrijke DA zijn gedaan, dat voor deze kopie data een back-up beleid is en eventuele analyses opnieuw gedaan kunnen worden. (aangenomen dat voor live data al een back-up policy bestaat en hierop gecontroleerd wordt)</i>

### 1.9 Voorbereidingstijd van gegevens uit andere bronnen

*Het verkrijgen van data uit andere (intern / extern) niet standaard bronnen vereist voorbereiding voor verwerking. Voor verwerking en interpretatie dient de ontvangen data geanalyseerd te worden voorafgaand aan verdere verwerking binnen DA en eventuele metadata gegevens doorgenomen te worden.*

<b>A</b>	<i>Focus: Plan structureel vooraf tijd in gericht op controle van ontvangen data, controleer op onregelmatigheden (ontbreken van bijv. datavelden, is de data omvang aannemelijk), controleer of verwerkingscapaciteit voldoende is. Weet van regelmatig benodigde en ontvangen data wat er aan data preparatie nodig is voor DA.</i>
<b>B</b>	<i>Focus: Weet structureel of data ook metadata bevat, probeer bij belangrijke data dit structureel te verkrijgen en desnoods structureel vast te leggen bij de interne data verzameling. Zorg voor het kunnen toetsen van DA uitkomsten, bekijk of er bijzondere standaardafwijkingen zijn. Verwerk / registreer metadata informatie in de analyse.</i>
<b>C</b>	<i>Focus: Zorg voor een vastgesteld proces van data verwerking bij DA waarbij een vast aantal data beoordelingsacties voor en na analyse aanwezig zijn.</i>

## Gegevenskwaliteit

### 2.1 Nauwkeurigheid

*De nauwkeurigheid geeft aan hoeveel de meetwaarde of informatie van de werkelijkheid af zit. Nauwkeurigheid is te bepalen door de meting te vergelijken met een meting waarvan de nauwkeurigheid bekend is.*

<b>A</b>	<i>Focus: Creëer bewustwording van data nauwkeurigheid, erken het belang ervan.</i>
<b>B</b>	<i>Focus: Leg bij de belangrijkste data vast hoe deze op nauwkeurigheid gecontroleerd kan worden en bij herhaalde analyses op dezelfde wijze kan worden uitgevoerd.</i>
<b>C</b>	<i>Focus: Breid de controle mogelijkheid uit naar zoveel als mogelijk en leg vast hoe nauwkeurighedscontroles structureel uitgevoerd moeten worden.</i>

### 2.2 Volledigheid

*Bepalen welke datavelden je nodig hebt voor de analyse. Daarnaast per dataveld controle of het daadwerkelijk gegevens bevat.*

<b>A</b>	<i>Focus:</i>
----------	---------------

	<i>Creëer bewustwording van data volledigheid, erken het belang ervan.</i>
<b>B</b>	<i>Focus: Leg bij de belangrijkste data vast hoe deze op volledigheid gecontroleerd kan worden en bij herhaalde analyses op dezelfde wijze kan worden uitgevoerd.</i>
<b>C</b>	<i>Focus: Bepaal en leg vast hoeveel data vereist is voor de analyses om voldoende kwaliteit te krijgen. Zorg voor overzicht van de verschillende formats waarin data is aangeleverd. Leg vast bij elke analyse waarop binnen de dataset gecontroleerd dient te worden om de volledigheid te toetsen</i>

### 2.3 Consistentie

*De gegevens mogen geen hiaten bevatten in het tijdspad van analyse (tijdconsistentie), geen data duplicaten en vrij van tegenstrijdigheden. Hebben meet sensoren bijvoorbeeld continue goed gefunctioneerd en data constant aangeleverd, zijn handmatige formulieren door iedereen op een zelfde wijze ingevuld.*

<b>A</b>	<i>Focus: Leg vast hoe data deduplicatie kan worden toegepast bij de analyses. Bij tijdsanalyses controleer via steekproef of op verwachte momenten ook een registratie is gedaan of zijn er blanco tijdsperiodes (achtereenvolgende momenten waarop data zou moeten zijn vastgelegd).</i>
<b>B</b>	<i>Focus: Zorg voor voldoende kennis en of technieken om hiaten in data consistentie te controleren, gebruik eventueel metadata om inconsistentie bij data verzameling vast te leggen.</i>
<b>C</b>	<i>Focus: Zorg voor voldoende kennis om data hiaten indien mogelijk aan te vullen, weet waar ontbrekende data gehaald kan worden. Als aanvulling niet kan, weet hoe de analyse afgerond kan worden met toevoeging van de juiste opmerking.</i>
<b>D</b>	<i>Focus: Zorg dat kennis en techniek aanwezig blijft om hiaten in data te detecteren.</i>

### 2.4 Geloofwaardigheid

*De mate waarin gegevens kenmerken hebben die in een specifieke gebruikcontext als waar en geloofwaardig (authentiek) kunnen worden beschouwd en de verzameling op een juiste wijze heeft plaatsgevonden.*

<b>A</b>	<i>Focus: Leg vast hoe bij een aantal (herhalende) analyses op geloofwaardigheid gecontroleerd kan worden. Is bijvoorbeeld de data gegarandeerd van de verwachte sensor of machine, is de data gelijk aan hetgeen verwacht kan worden kwantitatief, format, type, etc.</i>
<b>B</b>	<i>Focus: Zorg voor databron vermelding bij data verzameling, waar mogelijk een controle van tijd registraties periodiek uitvoeren. Vergaar inzicht in mogelijke data verzamel methode en welke toegepast worden. Creëer methode om herhaalde data geloofwaardigheid controle te kunnen uitvoeren.</i>
<b>C</b>	<i>Focus: Waar mogelijk analyse resultaten controleren op bijzondere uitkomsten en zo nodig terugkoppelen naar mogelijke oorzaken in de verzamelde data.</i>
<b>D</b>	<i>Focus: Verkrijg kennis/ervaring in het controleren van analyse resultaten op geloofwaardigheid en het herleiden naar verzamelde data. Wijzig indien nodig de methode van verzamelen.</i>



## 2.5 Actueel

*Bevat de data kenmerken van de juiste leeftijd binnen een specifieke analyse context.*

<b>A</b>	<i>Focus: Bekijk de aangeleverde data vooraf en bekijk of het datum – tijd informatie bezit wat beoordeeld kan worden. Weet vooraf aan de analyse welke periode van belang is.</i>
<b>B</b>	<i>Focus: Weet bij welke herhaalde analyses datum tijd van belang is, zorg voor kennis hoe dit gecontroleerd kan worden.</i>
<b>C</b>	<i>Focus: Zorg voor (automatische) structurele controles van veelgebruikte data op datum- en tijdwaardes zodat bij gebruik voor analyse de actualiteit kenbaar is.</i>
<b>D</b>	<i>Focus: Bewaak de structurele controles op data.</i>

## 2.6 Toegankelijkheid

*De mate waarin gegevens toegankelijk zijn in een specifieke gebruikscontext, met name door mensen die ondersteunende technologie of speciale configuratie nodig hebben vanwege een toegangsbeperking.*

<b>A</b>	<i>Focus: Zorg voor minimaal toegang tot een kopie van data bestanden of database. Pas desnoods de rechten aan zodat er in ieder geval deels toegang is tot data voor analyse.</i>
<b>B</b>	<i>Focus: Zorg voor een datakopie welke zich kan synchroniseren met de live data zodat actuele data-analyses (op live data) mogelijk is.</i>
<b>C</b>	<i>Focus: Zorg voor een directe data toegang en dat de rechten voldoende zijn zodat data-analyses op live data mogelijk is.</i>
<b>D</b>	<i>Focus: Zorg dat data analist naast bij bestaande ook bij nieuwe data bronnen voldoende rechten krijgen bij inrichting.</i>
<b>E</b>	<i>Focus: Controleer jaarlijks of de data analist nog steeds voldoende toegangsrechten heeft.</i>

## 2.7 Naleven

*De mate waarin gegevens kenmerken hebben die voldoen aan geldende normen, conventies of voorschriften en soortgelijke regels met betrekking tot gegevenskwaliteit in een specifieke gebruikscontext.*

<b>A</b>	<i>Focus: Achterhaal of er branchevoorschriften en of wettelijke regels bestaan voor data verwerking.</i>
<b>B</b>	<i>Focus: Verwerk de belangrijkste branchevoorschriften en of wettelijke regels bij de data verwerking</i>
<b>C</b>	<i>Focus: Verwerk de branchevoorschriften en of wettelijke regels bij de data verwerking</i>
<b>D</b>	<i>Focus: Zie toe op het hanteren van de branchevoorschriften en of wettelijke regels bij de data verwerking.</i>
<b>E</b>	<i>Focus: Stel interne regels op gericht op data en dataverwerking, gericht op toegang, privacy, etc.</i>

## 2.8 Vertrouwelijkheid

*Hoe vertrouwelijk is de data (analyse resultaten) binnen de organisatie. In welke mate is het toegankelijk en is deze interpreteerbaar zonder kennis van de context.*

<b>A</b>	<i>Focus: Probeer data toegang af te schermen voor die medewerkers die qua functie geen toegangsbelang hebben</i>
<b>B</b>	<i>Focus: Vergroot het vertrouwen in de data analist, beperkt zoveel mogelijk de toegang tot de data voor anderen.</i>
<b>C</b>	<i>Focus: Maximaliseer het vertrouwen in de data analist.</i>

## 2.9 Efficiëntie

*Het kunnen aanpassen van data zodat dit de prestatie van analyse verbetert bijvoorbeeld het verkleinen van de data omvang (deductie, compressie, ontdebellen)*

<b>A</b>	<i>Focus: Zorg dat er kennis is hoe belangrijke datasets aangepast kunnen worden zodat verwerking ervan voor analyse verbetert.</i>
<b>B</b>	<i>Focus: Zorg dat alle data indien nodig aangepast kan worden zodat verwerking ervan voor analyse mogelijk is.</i>
<b>C</b>	<i>Focus: Zorg dat alle datasets indien nodig aangepast kunnen worden zodat er een efficiëntere verwerking mogelijk is. Wellicht kan analyse software uitkomst bieden.</i>
<b>D</b>	<i>Focus: Beoordeel regelmatig of er efficiëntie behaald kan worden.</i>

## 2.10 Precisie

*In hoeverre bevat de data waardes welke, afhankelijk van de context op meerdere wijzen interpreteerbaar is.*

<b>A</b>	<i>Focus: Verdiep in het toepassen van verschillende analyse methodes</i>
<b>B</b>	<i>Focus: Verdiep in toepassen van verschillende analyse methodes en kennis van mogelijk verschillende uitkomsten.</i>
<b>C</b>	<i>Focus: Probeer verschillende analyse methodes op dezelfde data en beoordeel de uitkomsten op verschillen. Zoek per context naar goede (beste) analyse.</i>

## 2.11 Traceerbaarheid

*Kunnen we controleren op mutaties in de data nadat deze data is gebruikt voor analyse.*

<b>A</b>	<i>Focus: Weet hoe en waar je bij verschillende bestandtypes de meta data kan uitlezen. Achterhaal of op databases een traceerbaarheid mogelijk is.</i>
<b>B</b>	<i>Focus: Schakel aanwezige traceerbaarheidsmogelijkheden op databases in, controleer altijd op aanwezige metadata in bestanden.</i>
<b>C</b>	<i>Focus: Leg vast wanneer bij DA in het begeleidend schrijven informatie wordt vastgelegd over gemuteerde data wanneer dit mogelijk van belang is bij interpretatie van het resultaat.</i>

### 2.12 Begrijpelijkheid

*In hoeverre hebben gegevens kenmerken die het mogelijk maken om door gebruikers te lezen en te interpreteren, en worden uitgedrukt in geschikte talen, symbolen en eenheden in een specifieke gebruikscontext. Sommige informatie beschikt over metadata, een begeleidend schrijven.*

<b>A</b>	<i>Focus: Vergaar inzicht in het belang van de ondersteunende functie van metadata en achterhaal wat er reeds van aanwezig is.</i>
<b>B</b>	<i>Focus: Zorg voor interne bewustwording van het vastleggen van metadata bij belangrijke veel gebruikte data bronnen.</i>
<b>C</b>	<i>Focus: Weet wanneer achtergrond informatie ontbreekt bij data en leg dit vast in het begeleidend schrijven.</i>

### 2.13 Beschikbaarheid

*In hoeverre is de data direct gereed of is preparatie nodig om geanalyseerd te kunnen worden.*

<b>A</b>	<i>Focus: Verhoog kennis van aanwezige data typen binnen de organisatie, weet welke omgezet moeten worden voor efficiënte analyse en hoe deze geschikt gemaakt kunnen worden voor analyse.</i>
<b>B</b>	<i>Focus: Automatiseer het omzetten van data naar een geschikt formaat, daarmee wordt het altijd beschikbaar voor analyse.</i>
<b>C</b>	<i>Focus: Leg vast hoe de beschikbaarheid bij verschillende data wordt verkregen zodat dit herhaald kan worden.</i>

### 2.14 Transporteerbaar

*De mate waarin gegevens kenmerken hebben waarmee deze kunnen worden geïnstalleerd, vervangen of verplaatst van het ene systeem naar het andere, met behoud van de bestaande kwaliteit in een specifieke gebruikscontext.*

<b>A</b>	<i>Focus: Bij data omvang issues, verhoog capaciteit. Verhoog daarnaast het inzicht waarom het verplaatsen moeilijk is.</i>
<b>B</b>	<i>Focus: Bij technische beperking achterhaal wat nodig (kennis) is om dit wel te kunnen. Laat indien mogelijk consultant een zodanige aanpassing doen dat transporteerbaarheid vereenvoudigd wordt.</i>
<b>C</b>	<i>Focus: Leg vast hoe moeilijk verplaatsbare data toch verplaatst kan worden – instructie van consultant.</i>

### 2.15 Herstelbaarheid

*De mate waarin ruwe data kan worden hersteld naar de situatie waarop analyse heeft plaats gevonden en exact dezelfde analyse weer uitgevoerd kan worden.*

<b>A</b>	<i>Focus: Snapshot methodes zijn bij uitstek geschikt voor het terug halen van data, bepaal een acceptabele snapshot frequentie. Informeer bij database beheerder naar mogelijkheden.</i>
<b>B</b>	<i>Focus: Beschrijf een policy voor het gebruik van data en de herstelbaarheid zodat reproduceren van de analyse mogelijk is. Leg eventuele gebruikte query's vast.</i>

<b>C</b>	<i>Focus:</i> <i>Leg bij nieuwe analyses vast hoe om te gaan met herstelbaarheid.</i>
----------	------------------------------------------------------------------------------------------

## 2.16 Reproduceerbaar

*De mate waarin gegevens kenmerken hebben die het mogelijk maken om een bepaald niveau van bewerkingen en kwaliteit te behouden, zelfs in het geval van een storing, in een specifieke gebruikscontext (vastleggen van bijvoorbeeld gebruikte query, metadata beschrijving, eventuele gemaakte specifieke instellingen binnen een tool).*

<b>A</b>	<i>Focus:</i> <i>Verhoog analyse methode kennis bij de data analist</i>
<b>B</b>	<i>Focus:</i> <i>Bij belangrijke en of herhalende analyses vastleggen welke data is gebruikt en verwerkt in de analyse.</i>
<b>C</b>	<i>Focus:</i> <i>Stel regels op wat er bij analyse vastgelegd moet worden zodat reproduceerbaarheid geborgd wordt.</i>
<b>D</b>	<i>Focus:</i> <i>Voer audits uit gericht op vastlegging van toegepaste query's en metadata beschrijving.</i> <i>Hiermee bewaken van reproduceerbaarheid van processen.</i>

Bijlage 18, MKB kenmerken (Kartiwi & MacGregor (2007) i.c.m. dimensies.

	MKB eigenschappen	Int 1	Int 2	Int 3	Int 4	Int 5	Int 6	Int 7	Int 8	Int 9	Int 10	Int 11	Ext 1	Ext 2	Ext 3	Ext 4	Ext 5	Ext 6	Ext 7	Ext 8	
	Focus Areas																				
Gegevensbeschikbaarheid																					
1.1	bepierking in interne gegevensgebruik		v		v		v	v	v	v	v	v		v		v				v	
1.2	bepierking in externe gegevensgebruik	v		v			v	v		v	v	v		v		v		v		v	
1.3	Niet bestaande of ontbrekende gegevens		v	v	v	v	v	v	v			v		v		v		v		v	
1.4	Timing en synchronisatie van gegevens					v						v		v		v		v		v	
1.5	Gegevensfrequentie		v	v			v	v	v	v	v	v		v		v		v		v	
1.6	Gegevensformat, standaarden en specificaties											v		v		v		v		v	
1.7	Onzekerheid, betrouwbaarheid en nauwkeurigheid					v			v			v	v	v	v	v		v		v	
1.8	Toegang tot gegevens, opslag en verwerking				v		v	v				v		v		v		v		v	
1.9	Voorbereidingstijd van gegevens uit andere bronnen		v		v	v						v						v			
Gegevenskwaliteit (obv Data Quality-in-usemodel)																					
INHERENTE AFHANKELIJKE KENMERKEN																					
2.1	Nauwkeurigheid - Accuracy	v		v		v	v	v	v		v		v	v	v	v		v			
2.2	Volledigheid - Completeness		v	v	v	v	v	v	v			v	v	v	v	v		v		v	
2.3	Consistentie - Consistency		v	v	v	v	v	v	v		v		v	v	v	v		v		v	
2.4	Geloofwaardigheid - Credibility (Believability)				v		v	v	v				v	v	v	v		v			
2.5	Actueel - Currentness	v		v		v	v	v	v				v	v	v	v		v			
INHERENTE & SYSTEEM AFHANKELIJKE KENMERKEN																					
2.6	Toegankelijkheid – Accesibility			v	v	v	v	v				v		v		v					
2.7	Nakomen – Compliance		v	v	v	v	v	v	v					v		v					
2.8	Vertrouwelijkheid - Confidentiality – data analyst				v								v	v	v	v				v	
2.9	Efficiëntie - Efficiency	v	v	v		v	v	v	v		v	v		v		v		v			
2.10	Precisie - Precision			v		v	v	v	v					v		v					
2.11	Traceerbaarheid- Traceability			v		v	v	v	v			v		v		v				v	
2.12	Begrijpelijkheid - Understandability			v		v	v	v	v					v		v				v	
SYSTEEM AFHANKELIJKE KENMERKEN																					
2.13	Beschikbaarheid - Availability			v	v	v	v	v	v	v	v	v		v		v					
2.14	Transporteerbaar - Portability			v		v	v	v			v	v		v		v		v			
2.15	Herstelbaarheid – Recoverability (Raw data)	v			v		v	v	v			v	v		v		v				
2.16	Reproduceerbaar - Reproducible (Analyse proces)	v		v		v	v	v			v	v				v					

## Uitleg en argumentatie tussen de MKB kenmerken en de dimensies

ID	Unieke kenmerken van MKB, Kartiwi et al. (2007)	Beargumenteerd de relatie met Gegevenskwaliteitskenmerken van Merino et al. (2015)
INT 1	MKB's hebben een klein en gecentraliseerd beheer met een korte termijn perspectief.	Voor het aansturen van de org met korte termijn perspectief zijn frequente data-analyses nodig, soms is er maar een korte data verzamel periode: 1.2 als niet binnen de korte termijn van analyse de externe gegevens ontvangen / geleverd kunnen worden zullen deze niet meegenomen kunnen worden. 1.5 een hogere frequentie zal tot betere inzichten leiden, zeker van belang om binnen korte termijnen een goede data-analyses uit te voeren. 2.1 het actief en gericht verzamelen in een korte periode van data voor DA kan nauwkeurige data opleveren. 2.5 van verzamelde data binnen een recente en korte termijn weet je in belangrijke mate de actualiteit ervan. 2.9 Binnen een relatief korte termijn specifiek de focus op het gericht data verzamelen kan efficiënt uitgevoerd worden. 2.15 in geval van een storing zou (mede afhankelijk van backup retentietijd) een redelijke mate van herstelbaarheid aanwezig zijn zodat een analyse met gegevens van een korte termijn mogelijk is. 2.16 in geval van een storing zou (mede afhankelijk van backup retentietijd) een redelijke mate van reproduceerbaarheid aanwezig zijn zodat een analyse met gegevens van een korte termijn mogelijk is.
INT 2	MKB's hebben slechte management-vaardigheden.	Beperkte management vaardigheden heeft invloed op: 1.1 niet tijdig en onvoldoende data hebben bij DA 1.3 Verkeerde beslissingen nemen bij het verzamelen van data via sensoren. 1.5 bepalen van frequentie zal nauwkeurig moeten. 1.9 besluiten om onvoorbereid te starten van een DA 2.2 bij onvoldoende volledigheid al een DA starten. 2.3 idem 2.7 indien het niet toeziet op algemene geldende normen zal ook een eventueel DA uitkomst mogelijk niet compliant info bevatten. 2.9 onjuiste aansturing kan leiden tot inefficiëntie bij verzameling en minder data opleveren voor analyse. 2.16 bij het uitvoeren van DA door management en het verzuimen van de toegepaste methode en query's heeft invloed op reproduceerbaarheid.
INT 3	MKB's tonen een sterk verlangen naar onafhankelijkheid en vermijden zakelijke ondernemingen die hun onafhankelijkheid aantasten.	Verzamelen van interne data voor DA belangrijk. Via DA meer inzicht in interne processen. Dit verlaagt eventuele behoefte aan externe vergelijkbare proces informatie en geeft minder afhankelijkheid. 1.2 en 1.3 kan zijn dat er vanuit dit verlagen gekozen wordt voor minder behoefte aan externe gegevens. Bij gegevenskwaliteit geldt dit voor: 2.1 t/m 2.5, 2.6 t/m/ 2.12 en 2.13 t/m 2.16.
INT 4	MKB-eigenaren houden vaak informatie achter van collega's.	Het achterhouden van data waarbij deze niet meegenomen kan worden bij DA kan invloed hebben 1.1, 1.3, 1.8 en 1.9 en 2.2 (incomplete data set), 2.3 (hiaat in dataset), 2.6 (geen toegang tot delen in de dataset), 2.7 (verzwijgen van verplichtingen waaraan dient te worden voldaan bij data/DA), 2.8 vanwege vertrouwen niet delen van data, 2.13 door het achterhouden is data niet beschikbaar.
INT 5	Het besluitvormingsproces in MKB's is intuïtief, in plaats van op basis van gedetailleerde planning en uitputtende studie.	Als intuïtief beoordeeld wordt dat DA kan worden toegepast op de aanwezige data dan kan getwijfeld worden aan de kwaliteit van: 1.3, 1.4, 1.5, 1.7 en 1.9 2.1 t/m 2.5, 2.6 t/m/ 2.12 en 2.13 t/m 2.16.
INT 6	De MKB-eigenaar(s) heeft/hebben een sterke invloed in het besluitvormingsproces	MKB eigenaar veel invloed op de mate van volwassenheid van alle kenmerken: 1.1, 1.2, 1.3, 1.5 en 1.8 2.1 t/m 2.5, 2.6 t/m/ 2.12 en 2.13 t/m 2.16.

INT 7	Inbreuk op familiewaarden en zorgen in besluitvormingsprocessen.	Waar de familiewaarden veel invloed hebben op organisatie zal dit ook veel invloed op de volwassenheid van alle kenmerken hebben: 1.1, 1.2, 1.3, 1.5 en 1.8 2.1 t/m 2.5, 2.6 t/m/ 2.12 en 2.13 t/m 2.16.
INT 8	MKB's hebben een informele en ontoereikende planning en processen voor het bijhouden van records.	Het niet vastleggen en bijhouden van data raakt veel kenmerken en beïnvloed de gegevensbeschikbaarheid en gegevenskwaliteit: 1.1 beperkt bijhouden van dat levert beperking op bij interne gegevensgebruik. 1.3 informele processen geeft verhoogde kans op ontbrekende gegevens en op 1.5 wisselende registratie frequentie. Het beïnvloed 1.7. 2.1 onzekerheid over de nauwkeurigheid, 2.2 onvolledig, 2.3 niet consistent, 2.5 het niet kunnen beoordelen van data op actualiteit, 2.7 beperkt bijhouden van processen zegt mogelijk iets over de mate van Compliance, 2.9 beperkte proces standaardisatie geeft beperkte efficiëntie, 2.10 beperkte proces standaardisatie geeft beperkte precisie, 2.11 beperkte proces standaardisatie geeft beperkte Traceerbaarheid, 2.12 beperkte proces standaardisatie geeft beperkte vastlegging van begrijpelijkheid, 2.13 beperkte proces standaardisatie geeft beperkte standaardisatie van hoe data wordt vastgelegd. Daarmee kan dit ook van moment tot moment verschillen. Data is daarmee verschillend en niet direct beschikbaar voor DA verwerking, 2.15 Herstelbaarheid, de mate van ingerichte backup proces formaliteit bepaald de herstelbaarheid, 2.16 informele processen zorgt voor beperkte vastlegging hoe DA is uitgevoerd en vervolgens reproduceerbaar is.
<b>Functies met betrekking tot resource-acquisitie</b>		
INT 9	MKB's ondervinden moeilijkheden bij het verkrijgen van financiering en andere middelen, en als gevolg daarvan minder middelen hebben.	Beperkte middelen levert veelal beperkte data verzameling op. Los daarvan staat de kwaliteit van de verzamelde data wat voor DA gebruikt kan worden. 1.1 het beschikken over beperkte middelen zal een beperking geven op het gebruik van interne gegevens 1.2 beperking van externe gegevensgebruik 1.5 een hogere frequentie van data verzameling vereist meer opslagmiddelen. 2.13 Beschikbaarheid van externe data.  Overheid zet steeds meer data online toegankelijk voor MKB in de hoop dat deze dit steeds meer zal gebruiken bij DA. Vrij eenvoudig toegankelijke externe data. Gebruik van standaard analyses voor core processen vergt beperkte investering zodat DA deels mogelijk is.
INT 10	MKB's zijn terughoudender om uit te geven aan informatietechnologie en hebben daarom een beperkt gebruik van technologie.	Beperkte IT uitgaven levert beperkte digitale data verzameling op, handmatige lijsten blijft. Een direct gevolg is minder kwaliteit en kwantiteit van de verzamelde data. Beperkte uitgaven aan IT heeft weer invloed op de kwaliteit van het data-analyse proces en de uitkomsten. 1.1 het beschikken over beperkte middelen zal een beperking geven op het gebruik van interne gegevens 1.2 beperking van externe gegevensgebruik 1.5 een hogere frequentie van data verzameling vereist meer opslagmiddelen 2.1 minder data betekend minder nauwkeurige analyses, 2.3 beperkte IT gebruik kan leiden tot minder consistentie van data. 2.9 verzamelen van data via andere methodes dan IT zijn veelal minder efficiënt. 2.13 beperkte IT middelen kan leiden tot beperkte beschikbaarheid van data, als data op een lokale pc staat en deze niet beschikbaar / benaderbaar is. 2.14 Wanneer er beperkt in IT middelen wordt geïnvesteerd, zorg dat noodzakelijke data (transporteerbaar) geoutsourcet kan worden en eventueel daar voor DA ingezet. Bij minimale IT inzet komt het gevaar van 2.15 herstelbaarheid en 2.16 reproduceerbaarheid al snel in de verdrukking.

<b>INT 11</b>	MKB's hebben een gebrek aan technische kennis en specialistische personeel en bieden weinig IT-training voor personeel.	Het gebrek aan kennis raakt elke item van gegevensbeschikbaarheid 1.1 t/m 1.9 2.2 bij gebrek aan kennis is het bepalen van volledigheid lastig. 2.6 kennis tekort kan leiden tot het niet weten welke data er allemaal is en hoe deze toegankelijk is. 2.9 beperkte data/IT kennis levert mogelijk beperkingen op bij het efficiënt maken van data voor DA. 2.11 beperkte data/IT kennis levert mogelijk beperkingen op bij data mutatie logging beheren. 2.13 kennis tekort kan leiden tot het niet weten hoe data benaderbaar is, welke rechten er nodig zijn zodat deze beschikbaar is. 2.14 kennis tekort kan leiden tot verkeerd verplaatsen van data waardoor deze voor DA kwaliteit verliest. 2.15 Als kennis over hoe data herstellen beperkt is heeft dit invloed op DA bij een storing. 2.16 Beperkte IT kennis kan leiden tot foutieve registratie van informatie welke nodig is om DA te reproduceren.
<b>Kenmerken met betrekking tot producten/diensten en markten</b>		
<b>EXT 1</b>	MKB's hebben een smal product-/dienstenassortiment	Met dit MKB kenmerk kan het MKB zich juist onderscheiden van grotere bedrijven door zorgvuldig data verzameling voor DA toe te passen. 1.7 om specifieke kwalitatieve informatie te creëren is dit nodig. 2.1 nauwkeurige data van productie proces, product eigenschappen en of specifieke markt behoeften zijn belangrijk om zicht te kunnen onderscheiden met DA. 2.2, 2.3, 2.4 en 2.5 zijn in dat zelfde kader van belang. 2.8 als de data voor iedereen toegankelijk is kan ook de concurrent er mogelijk bij komen, zorg voor toegangscontrole om toegang tot data te krijgen. .
<b>EXT 2</b>	MKB's hebben een beperkt marktaandeel (vaak beperkt tot een nichemarkt) en vertrouwen daarom op zwaar op weinig klanten.	Een beperkt marktaandeel met een paar klanten kan zeer specifieke (regionaal of nichemarkt gerichte) kennis opleveren op basis van verzamelde data. De markt en het aantal klanten kent geen directe relatie met de Data Analytics processen en de volwassenheid daarvan binnen een MKB organisatie.
<b>EXT 3</b>	MKB's zijn productgericht, terwijl grote bedrijven Klantgericht.	Ook met dit MKB kenmerk kan het zich juist onderscheiden van grotere bedrijven door zorgvuldig data verzameling voor DA toe te passen. 1.7 om specifieke kwalitatieve informatie te creëren is dit nodig. 2.1 nauwkeurige data van productie proces en product eigenschappen zijn belangrijk om zicht te kunnen onderscheiden met DA. 2.2, 2.3, 2.4 en 2.5 zijn in dat zelfde kader van belang. 2.8 als de data voor iedereen toegankelijk is kan ook de concurrent er mogelijk bij komen, zorg voor gepaste toegangscontrole om toegang tot data te krijgen.
<b>EXT 4</b>	MKB's zijn niet geïnteresseerd in grote marktaandelen	Het hebben van veel of weinig markt aandeel zegt niks over het moeten nemen van de juiste beslissingen. DA ondersteund organisaties bij het nemen van beslissingen door informatie aan te leveren. Mogelijk dat de dimensie gericht op externe data als weinig van belang zal worden beoordeeld.
<b>EXT 5</b>	MKB's kunnen niet concurreren met hun grotere tegenhangers.	Geen bijzondere gegevensbeschikbaarheid of gegevenskwaliteitskenmerk wat hierbij in vergelijking tot de andere een belangrijke mate een relatie heeft.
<b>Kenmerken met betrekking tot het nemen van risico's en het omgaan met onzekerheid</b>		
<b>EXT 6</b>	MKB's hebben minder controle over hun externe omgeving dan grotere bedrijven, en worden daarom geconfronteerd met meer onzekerheid	Externe gegevensbeschikbaarheid en gegevenskwaliteit zijn belangrijk. 1.2, 1.9 2.1 t/m 2.5, 2.9 of de data direct verwerkt kan worden en 2.14 of de externe data zonder kwaliteitsverlies ontvangen kan worden.
<b>EXT 7</b>	MKB's lopen meer risico's dan grote bedrijven, omdat het aantal mislukkingen van MKB's hoger zijn.	Geen bijzondere gegevensbeschikbaarheid of gegevenskwaliteitskenmerk wat hierbij in vergelijking tot de andere een belangrijke mate een relatie heeft. .
<b>EXT 8</b>	MKB's zijn terughoudender om risico's te nemen	Het verzorgen dat er kwalitatief goede data verzamelt kan worden kent op zichzelf weinig risico. Terughoudendheid bij het verzamelen van data kan de beschikbaarheid en datakwaliteit beïnvloeden waaronder: 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7 en 1.8 2.2 datasets niet volledig, 2.3 inconsistente data, 2.8 dermate vertrouwelijk dat data niet vrijgegeven wordt voor DA, 2.11 bij



		risico's mijdende houding past het hebben van mutatie logging wat weer prima is bij traceerbaarheid. 2.12 bij opgeslagen data kan specifieke coderingswaardes worden gebruikt wat zonder achtergrond kennis niet begrijpbaar is.
--	--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------